

**UNDERSTANDING DISTORTION IN SILICON-GERMANIUM  
TRANSISTORS, AND ITS APPLICATION TO RF CIRCUITS**

A Thesis  
Presented to  
The Academic Faculty

by

SACHIN SETH

In Partial Fulfillment  
of the Requirements for the Degree  
Master of Science in the  
School of Electrical and Computer Engineering

Georgia Institute of Technology  
December 2009

**COPYRIGHT 2009 BY SACHIN SETH**

# **UNDERSTANDING DISTORTION IN SILICON-GERMANIUM TRANSISTORS, AND ITS APPLICATION TO RF CIRCUITS**

Approved by:

Dr. John D. Cressler, Advisor  
School of Electrical and Computer Engineering  
*Georgia Institute of Technology*

Dr. Joy Laskar  
School of Electrical and Computer Engineering  
*Georgia Institute of Technology*

Dr. Shyh-Chiang Shen  
School of Electrical and Computer Engineering  
*Georgia Institute of Technology*

Date Approved: 11/12/2009

## ACKNOWLEDGEMENTS

I wish to start by thanking Prof. John D. Cressler for giving me the opportunity to work under his guidance. Before I joined the SiGe research team, I knew he was good. Having worked with him, I can safely claim that he is the best advisor, and the best teacher, that any student could ever have. He continues to amaze me even after two years, and I hope to learn a lot more lessons from him, both technical and non-technical, during my graduate studies under him.

I would like to express my sincerest gratitude to my father, mother and sister for being supportive of me continuously during good and bad times. Not being close to them saddens me, and I can't thank them all enough for making graduate school easier for me, every day of the week.

I would also like to thank my colleagues at the SiGe research team, especially Curtis, for teaching me the ropes of the trade and teaching me to analyze a difficult concept such as linearity both experimentally and mathematically. I also would like to thank John and Stan for motivating me and engaging me in intelligent technical discussions.

Lastly, all acknowledgements would mean nothing if it were not for my fiancé, best friend, confidante and love of my life Manasi. Her motivational speeches and her abilities to always bring out the best in me are what made me attend graduate school at GT. A very fine team we make, and I am very blessed to be able to spend the rest of my life with her! Love you, Manasi.

# TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
LIST OF FIGURES	vi
LIST OF ABBREVIATIONS	viii
SUMMARY	ix
 <u>CHAPTER</u>	
<b>1 INTRODUCTION.....</b>	<b>1</b>
1.1 Motivation.....	1
1.2 SiGe BiCMOS Overview.....	2
1.3 Introduction to Distortion/Linearity.....	5
1.3.1 Gain Compression.....	6
1.3.2 Intermodulation Distortion.....	7
1.4 Summary.....	10
<b>2 USING VOLTERRA SERIES TO UNDERSTAND DISTORTION.....</b>	<b>11</b>
2.1 What is Volterra Series?.....	11
2.2 Contributors of Distortion in SiGe HBT.....	12
2.2.1 The $I_{CE}$ Exponential Nonlinearity.....	13
2.2.2 The $I_{BE}$ Exponential Nonlinearity.....	14
2.2.3 The $C_{BE}$ and $C_{BC}$ Nonlinearities.....	15
2.3 Volterra Series Applied to Distortion.....	16
2.4 Summary.....	22

<b>3</b>	<b>DISTORTION IN COMMON BASE CONFIGURATION.....</b>	<b>23</b>
3.1	Comparison of CE and CB Linearity.....	23
3.2	Compact Model Validation for CB linearity.....	25
3.3	Calculation of CB linearity using Volterra Series.....	28
3.4	Understanding Role of Load Impedance on Distortion.....	33
3.5	Summary.....	36
<b>4</b>	<b>COMPARING LINEARITY OF COMPLEMENTARY SIGE HBTs.....</b>	<b>37</b>
4.1	Complementary Technology Overview.....	37
4.2	Two-tone Power Sweep .....	40
4.3	Effect of Collector bias on Distortion.....	42
4.4	Comparison of <i>nnp</i> and <i>pnp</i> Distortion.....	44
4.5	Effect of Scaling Frequency on Linearity.....	47
4.6	Understanding Complementary Linearity Response.....	48
4.7	Summary.....	49
<b>5</b>	<b>CONCLUSIONS.....</b>	<b>51</b>
5.1	Conclusions.....	51
5.2	Future Directions.....	52
	<b>REFERENCES.....</b>	<b>54</b>

# LIST OF FIGURES

	Page
Figure 1: The evolution of $f_T$ and $f_{MAX}$ for SiGe HBTs across many generations.....	2
Figure 2: Vertical SIMS profile showing doping concentrations and Ge profile inside a first generation SiGe HBT.....	3
Figure 3: A schematic cross section of a third generation BiCMOS SiGe HBT.....	4
Figure 4: Amplifier outputs for 2 closely spaced input frequencies.....	8
Figure 5: Fundamental ( $1^{st}$ ) and $3^{rd}$ order intermod ( $3^{rd}$ ) outputs vs. increasing input signal power $P_{IN}$ .....	9
Figure 6: Equivalent circuit for $nnp$ SiGe HBT delineating dominant sources of nonlinearity.....	13
Figure 7: Algorithmic flow of the Volterra Series.....	17
Figure 8: Simplified equivalent $\pi$ -circuit of a transistor incorporating the four major nonlinearity sources.....	17
Figure 9: Second order transistor equivalent circuit for use in CMNA.....	19
Figure 10: The fundamental ( $P_{out-1^{st}}$ ) and $3^{rd}$ order intermodulation term ( $P_{out-3^{rd}}$ ) vs. Input power for CE and CB configurations.....	24
Figure 11: Comparison of CE and CB linearity across bias.....	24
Figure 12: Measured data to model (VBIC) comparison for the CB topologies at varying load impedances.....	26
Figure 13: Simplified network used to solve the (a) first-order and (b) second-order transfer functions for a CB SiGe HBT. The third-order network is the same as (b) with “ $i_{NL2}$ ” nonlinear current sources replaced by “ $i_{NL3}$ ” sources.....	29
Figure 14: $3^{rd}$ order virtual nonlinear current sources for nonlinear $g_m$ and $C_{BC}$ , comparing the full Volterra expressions with the simplified expressions 3.5 (5) and 3.6 (6).....	31
Figure 15: CB linearity calculated using the VS approach, and compared with simulation from VBIC model.....	33

Figure 16: Calculated CB linearity performance over bias for a SiGe HBT at different $Z_L$ . Top curve shows only IIP3 vs. $I_C$ with varying $Z_L$ . Bottom curve shows the virtual nonlinearity current sources with changing $Z_L$ . Only $i_{NL3Cbc}$ changes with $Z_L$ .....	34
Figure 17: Cross-section of <i>nnp</i> and <i>pnnp</i> SiGe HBTs developed by NSC used in this study.....	38
Figure 18: Gummel plots for similar sized <i>nnp</i> and <i>pnnp</i> SiGe HBTs.....	39
Figure 19: Cut-off frequencies $f_T$ normalized to peak $f_T$ of transistors vs current density $J_C$ .....	39
Figure 20: Fundamental ( $P_{FUND}$ ) and third order intermod term ( $P_{3rd}$ ) as a function of input power $P_{IN}$ for <i>nnp</i> SiGe HBT.....	40
Figure 21: Fundamental ( $P_{FUND}$ ) and third order intermod term ( $P_{3rd}$ ) as a function of input power $P_{IN}$ for <i>pnnp</i> SiGe HBT.....	41
Figure 22: IIP3 and Gain vs. current density $J_C$ with increasing $V_{CE}$ for standard sized <i>nnp</i> SiGe HBT.....	42
Figure 23: IIP3 and Gain vs. current density $J_C$ with increasing $V_{CE}$ for standard sized <i>pnnp</i> SiGe HBT.....	44
Figure 24: IIP3 and gain vs. current density ( $J_C$ ) with fixed bias for a standard sized <i>nnp</i> and <i>pnnp</i> SiGe HBT.....	45
Figure 25: Effect of scaling input frequency on IIP3 for <i>nnp</i> and <i>pnnp</i> SiGe HBTs in the high and low current density region.....	47
Figure 26: $i_{NL3gm}$ for similar sized <i>nnp</i> and <i>pnnp</i> SiGe HBT at similar bias.....	49

## LIST OF ABBREVIATIONS

BV	Breakdown Voltage
CB	Common Base
$C_{BE}$	Base-Emitter Capacitance
$C_{BC}$	Base-Collector Capacitance
CE	Common Emitter
CMNA	Compact Modified Nodal Analysis
HB	Harmonic Balance
HBT	Heterojunction Bipolar Transistor
IIP3	Input Third Order Intercept Point
IMD3	Third Order Intermodulation Distortion
IP3	Third Order Intercept PointSymbol-2 Means
OIP3	Output Third Order Intercept Point
VBIC	Vertical Bipolar Inter-Company
VS	Volterra Series



## SUMMARY

This thesis presents a study of distortion phenomenon in state of the art Silicon-Germanium (SiGe) Heterojunction Bipolar Transistor based circuits. Distortion/linearity is a key spec in designing high performance RF circuits, and both analytical methods and experimental results are discussed to ease the understanding of this tough concept.

Chapter 1 presents an overview of the SiGe BiCMOS technology. A background for the issues that distortion creates at a circuit and system level is given.

Chapter 2 presents a discussion of the Volterra Series mathematical analysis. It is explained how this concept can be applied to study and understand the problem of distortion, thus laying the foundations upon which chapter 3 and 4 are built.

Chapter 3 discusses the linearity response of SiGe HBTs in a Common Base configuration. The devices used in this study are from IBM's 3<sup>rd</sup> generation platform. The problem of distortion is tackled mathematically by employing Volterra Series, and analytical insights are given into transistor linearity. Compact expressions to predict bias levels for achieving maximum linearity in circuits are derived. The results in this chapter were published in *IEEE Bipolar/BiCMOS Circuits and Technology Meeting 2008* [18], and some results are pending journal publication in *IEEE Transactions of Microwave Theory and Techniques* [19].

Chapter 4 discusses the differences in linearity for electrically matched (similar  $f_T$  and BV) *nnp* and *pnp* SiGe HBTs. Experimental linearity studies are performed across a range of bias and load-matching conditions, and reasons for observed linearity

differences based in device physics are furnished. The results in this chapter were presented at *IEEE Bipolar/BiCMOS Circuits and Technology Meeting 2009* [27],

Chapter 5 summarizes the work presented, and lays down the future directions in which this work could be extended.

# CHAPTER 1

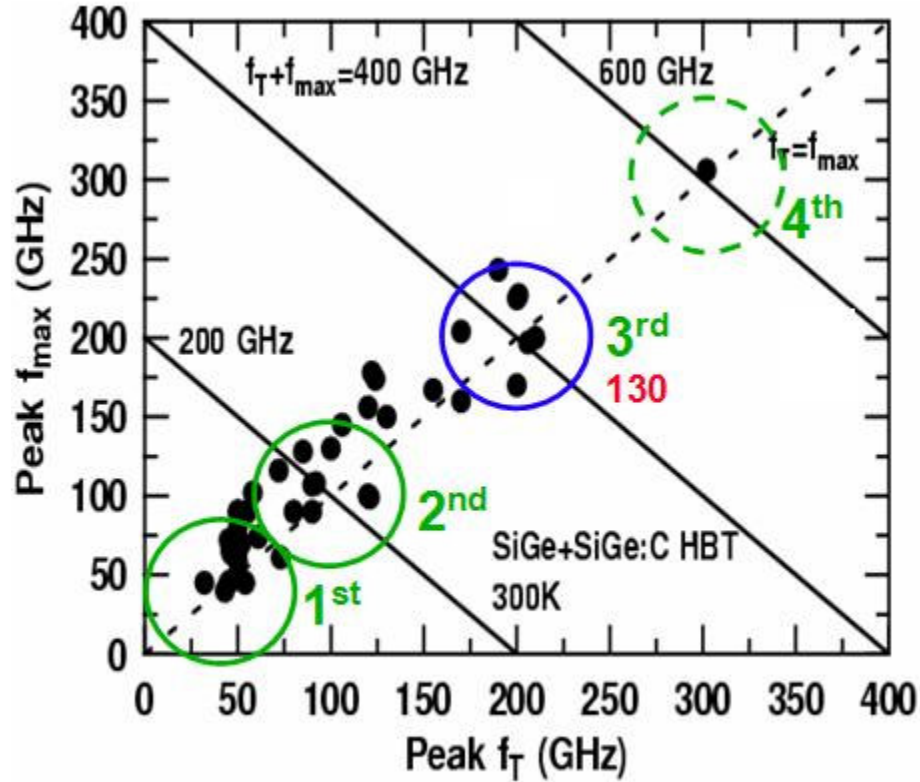
## INTRODUCTION

### *1.1 Motivation*

Silicon Germanium (SiGe) BiCMOS technologies continue to receive increased interest and deployment in the Radio Frequency (RF) circuit community because of their ever increasing performance metrics. Currently, with a peak unity gain frequency ( $f_T$ ) and maximum oscillation frequency ( $f_{MAX}$ ) approaching 500 GHz (Figure 1), these devices are ideal for use in high performance circuits like RF low noise amplifiers (LNA), power amplifiers (PA), and other sub-circuits in the transmit-receive chain like mixers, voltage controlled oscillators (VCO) etc [1], [2]. Such aggressive devices have been enabled due to continuous innovation in process integration and technology scaling. As a result, SiGe based circuits and systems today display similar performance as their III-V counterparts like GaAs and GaN based circuits, while maintaining the low cost advantages of CMOS technologies. With a rapidly increasing consumer base in RF technologies, usage of low-cost high-performance SiGe technologies is a win-win. For that reason, it becomes imperative for us to study SiGe technologies, with an emphasis on device metrics that directly map into highly aggressive RF circuits [3].

The enabler of these metrics in the SiGe technologies has been the incorporation of germanium in the base of a standard Si BJT. Using band-gap engineering, SiGe enjoys improvements in metrics like  $\beta$ ,  $V_A$  and  $f_T$  over the standard Si BJT counterparts [1]. And the ease of integration of Germanium in Base with standard Si processes permits superior levels of system complexity, leveraging the benefits of the best of breed Si CMOS technologies for potent use in mixed-signal application domain. The ability to fabricate

high performance SiGe devices on the same wafer as CMOS lends us the ability to develop strong system on chip architecture that offers low cost, reduced chip count, ease of packaging, low complexity all at much reduced costs. [1]

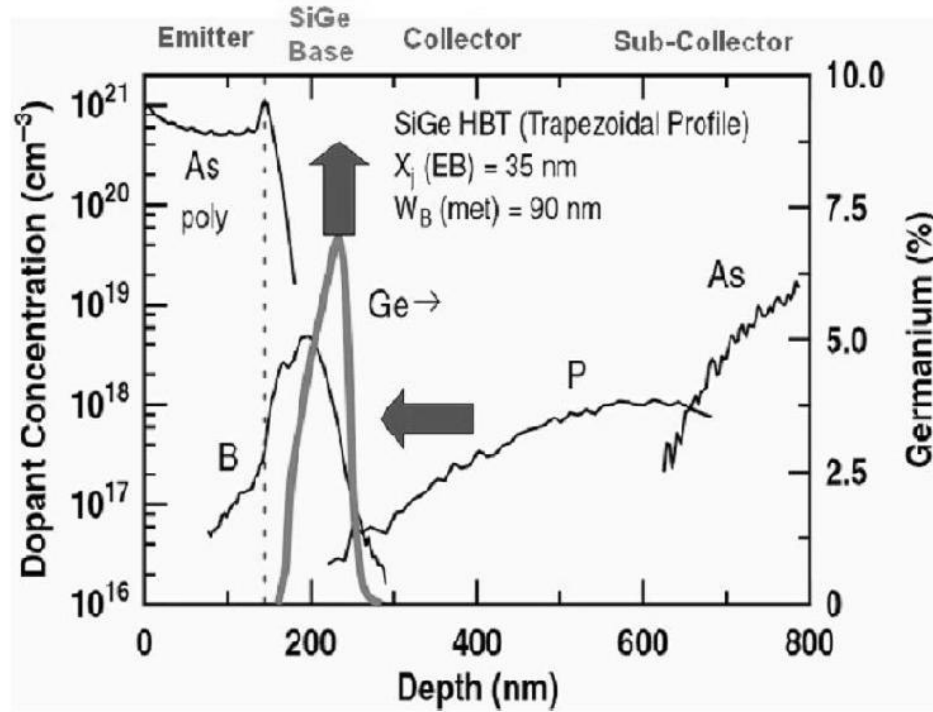


**Figure 1:** The evolution of  $f_T$  and  $f_{MAX}$  for SiGe HBTs across many generations [1].

## 1.2 SiGe BiCMOS Overview

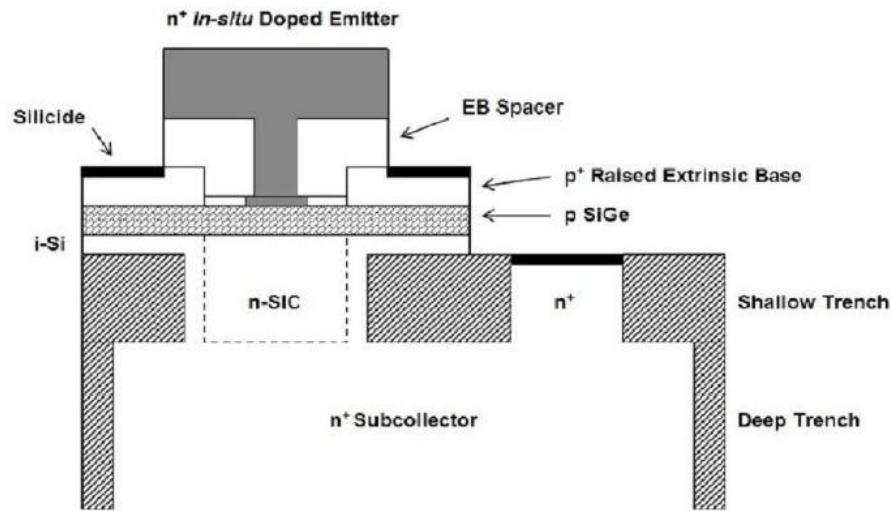
As mentioned previously, the one difference between a Si BJT and a SiGe HBT is the inclusion of the graded germanium in the active base region of the Si BJT. In  $n_{pn}$  HBT, this would imply adding a layer of compositionally graded SiGe alloy in the boron doped epitaxial layer. Figure 2 is the SIMS profile for a first generation SiGe HBT. In this figure, the placement of the SiGe alloy in the base region can be observed. This germanium layer is usually grown epitaxially using Ultra High Vacuum/Chemical Vapor

Deposition (UHV/CVD) technique [4]. This step does not tamper with the overall thermal budget of the fabrication process. This step is an easy add-on to the standard Si CMOS process flow, and does not affect the HBT performance, the CMOS yield and overall throughput [1]. This epitaxial layer is what enables bandgap engineering, which leads to high performance devices while still maintaining compatibility of process with standard Si CMOS.



**Figure 2:** Vertical SIMS profile showing doping concentrations and Ge profile inside a first generation SiGe HBT [1].

Several generation of SiGe HBTs exist in production, and are used in a host of application including cellular phones, wireless telemetry, radar systems and satellite communications. The devices studied in this thesis are the third-generation *npn* SiGe HBTs from IBM, as well as complementary SiGe HBTs from National Semiconductor (NSC).



**Figure 3:** A schematic cross section of a third generation BiCMOS SiGe HBT [2].

A cross section of the third generation devices from IBM is shown figure 3. Due to a number of advantages, a vertical self-aligned scheme is used. Some of these include reduced parasitics that give the device aggressive  $f_{MAX}$  metrics. A thin base region translates directly into a faster transit time for the electron from emitter to collector, lending itself to a faster  $f_T$  performance. To prevent the base region from diffusing out and thus compromising its thinness, it is imperative to employ a UHV/CVD epitaxial step that does not tax the thermal budget of the process flow too much. Also carbon may be using in addition to boron to prevent the out diffusion of the base in other process steps. Thus, with UHV/CVD and carbon in base, a narrow base profile is maintained. The technology also incorporated Deep Trench Isolation (DTI) and Shallow Trench Isolation (STI). A Selectively-Implanted Collector (SIC) enables IBM to manufacture the HBTs with varying speeds and breakdowns.

### ***1.3 Introduction to Distortion/Linearity***

As established before, SiGe technologies have found widespread use in wireless LAN, cellular telephony, satellite systems etc. Unlike wired communication, these types of communications all rely on a common transmission medium. The available spectrum is unfortunately limited, and hence interference with other transmissions has to be avoided at all costs. Various transistor, circuit and system level techniques have to be applied to ensure adequate circuit performance within the limited spectrum allotted for its operation [5].

Keeping above constraints in mind, it behooves the transmission system to

1. Utilize the frequency spectrum allotted to it efficiently, without causing interference.
2. Be able to work with signal ranging from very weak (close to “Noise Floor”) to very strong (operating in the large signal regime) without loss of signal integrity

However, for a system to live out these demands is not easy. Firstly, amplifiers have a tendency to generate spurious frequencies at their output that could interfere with other bands. This phenomenon is called “Intermodulation Distortion”. Secondly, circuits have a very limited Dynamic Range (DR). DR is the ratio of the “workable-signal” for an amplifier – the ratio between the maximum and minimum input signals for which a circuit can give acceptable signal quality. The lower end of the DR is the Noise Floor. The upper end is signified by the P1dB point – the input power level where the output gain has been clipped by 1 dB. This phenomenon of gain compression at high input power levels at amplifiers restricts them from carrying out point # 2 above. Gain compression is given rise due to large signal nonlinearities. Hence, non-linearities are of

2 types- “Intermodulation Distortion”, and “Gain Compression”. They are discussed one by one.

Before we start the next section, the reader should be made aware of the fact that the terms “nonlinearity” and “distortion” will be interchangeably used within this thesis. In an RF context, both terms signify the same phenomenon, that of a circuit or a system not giving outputs linearly with inputs.

### ***1.3.1 Gain Compression***

Assume an amplifier is fed with a single excitation frequency  $\omega_1$ . The amplified output should contain the same frequency  $\omega_1$  but magnified in its signal amplitude. In an ideal amplifier, the input  $x(t)$  and output  $y(t)$  should follow the relationship:

$$y(t) = k_1 x(t) \tag{1.1}$$

Where  $k_1$  = gain of the amplifier. Due to inherent non-linearities in the transistor, however, the input-output relation is governed by the following expression:

$$y(t) = k_1 x(t) + k_2 x^2(t) + k_3 x^3(t) + \dots \tag{1.2}$$

As a result, for an input  $x(t) = A \cos(\omega_1 t)$ , the output takes the shape



$$\begin{aligned}
y(t) = & \frac{k_2 A^2}{2} && \text{DC shift} \\
& + \left( k_1 A + \frac{3k_3 A^3}{4} \right) \cos \omega t && \text{Fundamental term} \\
& + \frac{k_2 A^2}{2} \cos 2\omega t && \text{Second harmonic} \\
& + \dots && (1.3)
\end{aligned}$$

Thus the output is not merely an amplified input frequency, but also a DC offset term, as well as higher order harmonics. The gain of this amplifier is not merely “ $k_I$ ” anymore, but a complex term  $\left( k_1 + \frac{3k_3 A^2}{4} \right)$ .

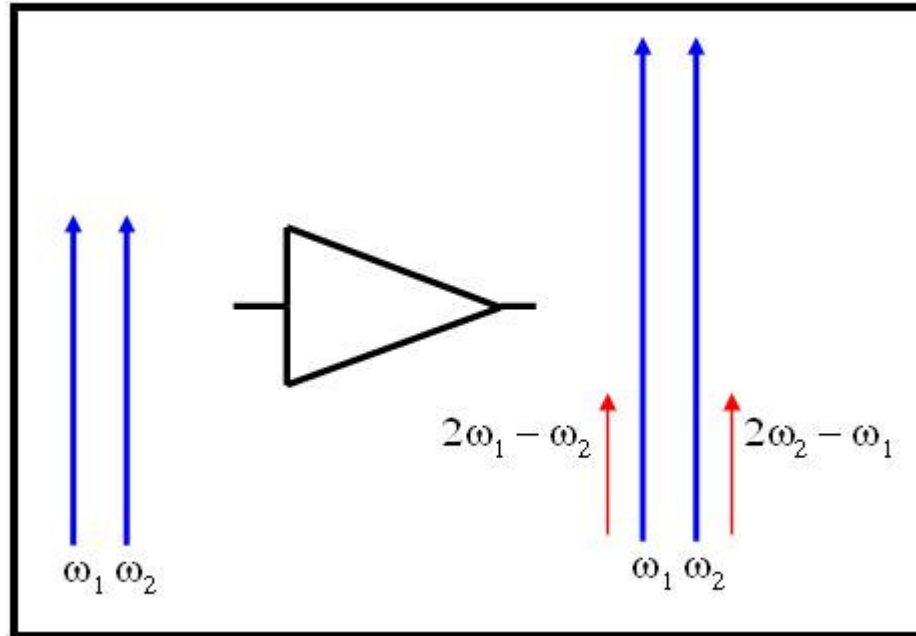
For small input signal A, the second term in the fundamental output expression can be neglected. But with increasing input signal A,  $3k_3 A^3/4$  becomes equal to, and then larger than and actual output  $k_I A$ . Usually  $k_3 < 0$ , which translated into diminishing gain values at sufficiently high input signal A to an amplifier. The gain thus “compresses” with increasing input. This is a fundamental manifestation of nonlinearity. This leads to a degraded Signal to Noise ratio (SNR) in RF circuits, as well as causes Bit Error Rates BER in digital circuits.

### ***1.3.2 Intermodulation Distortion***

Assume two closely spaced tones  $\omega_1$  and  $\omega_2$  of similar amplitude being fed at the input of an amplifier. The input sinusoid can be defined  $x(t) = A \cos(\omega_1 t) + A \cos(\omega_2 t)$ . Using the input output relation described in equation 1.2, the output partially is:

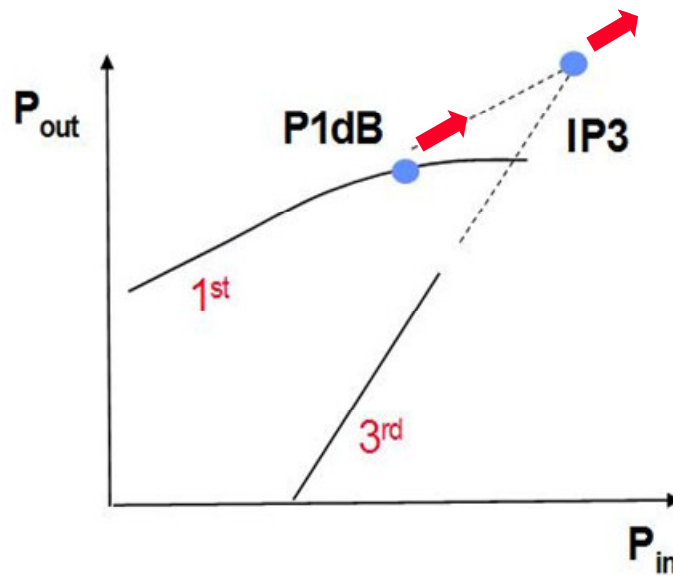
$$\begin{aligned}
y(t) = & \left( k_1 A + \frac{3k_3 A^3}{4} + \frac{3k_3 A^3}{2} \right) \cos \omega_{1,2} t + \dots && \text{Fundamental term} \\
& + \frac{3k_3 A^3}{4} \cos(2\omega_{1,2} - \omega_{2,1}) t + \dots && \text{3rd order Intermodulation} \\
& + \dots && (1.4)
\end{aligned}$$

From the above output expression 1.4, DC offset and harmonics have been neglected to maintain compactness. The intent is to show the reader the appearance of  $2\omega_1 - \omega_2$  and  $2\omega_2 - \omega_1$  frequencies at the output. These are called the third order intermodulation terms. The problem with these two spurious signals is that they are close to the frequencies of interest  $\omega_1$  &  $\omega_2$ , and can't be filtered out. These have the potential to fall in pass-bands of other communications in the vicinity and cause interferences. A pictorial description of the concept discussed above is shown in Figure 4 below.



**Figure 4:** Amplifier outputs for 2 closely spaced input frequencies.

A diagram that can tie both kinds of nonlinearity discussed above is shown in Figure 5. For small but increasing input signal with amplitude  $A$  the fundamental output grows linearly with it (as also derived in expression 1.4, neglecting higher order terms for small input), whereas the 3<sup>rd</sup> order intermod term increases as a cube. Plotted on a decibel power scale, the fundamental grows with a 1:1 slope with input signal power, and the 3<sup>rd</sup> order intermod grows with a 3:1 slope (due to its cubic dependency on input signal). This phenomenon is shown in the figure below.



**Figure 5:** Fundamental (1<sup>st</sup>) and 3<sup>rd</sup> order intermod (3<sup>rd</sup>) outputs vs. increasing input signal power  $P_{IN}$ .

The reader might wonder why only 2 curves are shown here, instead of the 4 outputs of an amplifier as shown too in Figure 4. The reason for that is that the signal amplitudes for the two fundamental are similar, as is the case for the 3<sup>rd</sup> order intermod terms (as seen from equation 1.4). Therefore to enable clarity of display, we are showing only one fundamental output and just one intermod term. P1dB is the point where the fundamental stops growing with a 1:1 relationship between output and input, and gain compression

sets in (described in section 1.3). The hypothetical extensions and intersection of the fundamental output curve and the 3<sup>rd</sup> order intermod output curve gives us what is called the 3<sup>rd</sup> order Intercept Point (IP3), and is the benchmark by which amplifier linearity is quantified. The x-coordinate of IP3 is known as the Input Third-Order Intercept Point (IIP3), and the y-coordinate of IP3 is known as the Output Third-Order Intercept Point (OIP3). IIP3 and OIP3 are very key figures of merit, and will be used extensively in later chapter to quantify linearity.

## ***1.4 Summary***

In this chapter, we discussed the advantage that SiGe based technologies hold over other technologies. We discussed the issues that make SiGe a contending candidate for use in wireless and RF systems. The structure of some devices studied in thesis was also outlined. We discussed the needs of emerging wireless systems, and discussed briefly some issues that beset those systems. Those will be studied in greater detail in the following chapters.

## CHAPTER 2

### USING VOLTERRA SERIES TO UNDERSTAND DISTORTION

#### *2.1 What is Volterra Series?*

The Volterra Series (VS) analysis is a mathematical tool that can be used to model weak nonlinearities in a system displaying distortion [6]-[8]. This technique dates back to the 1960s, and utilizes the theories developed by an Italian mathematician, Vito Volterra, to develop input/output expression for nonlinear system components. This tool can be used in complement with another technique for distortion analysis – namely the “Harmonic Balance” (HB) approach [9]. The advantages of VS over HB approach are that it is a set of algebraic expressions, the calculations of which are easily carried out resulting in rapid convergence in present day circuit simulators. Convergence is typically a problem that the HB approach has to cope with. Secondly, VS gives us an unbeatable advantage over HB in the analysis of nonlinearity for a circuit – that of decoupling the dominant nonlinearity sources, thus enabling their mitigation in a piecewise manner [10].

The mathematics of VS approach is undoubtedly cumbersome. The attempt in this thesis will be to discuss as much mathematics as is sufficient, without entangling the reader in a web of equations. Fortunately, delving deeply into the mathematics of the VS is not required to make good use of this technique in circuit analysis. It is more important to accurately model device and circuit level nonlinearities while understanding the prime limitation of the VS approach - it cannot be used to model strong nonlinearities.

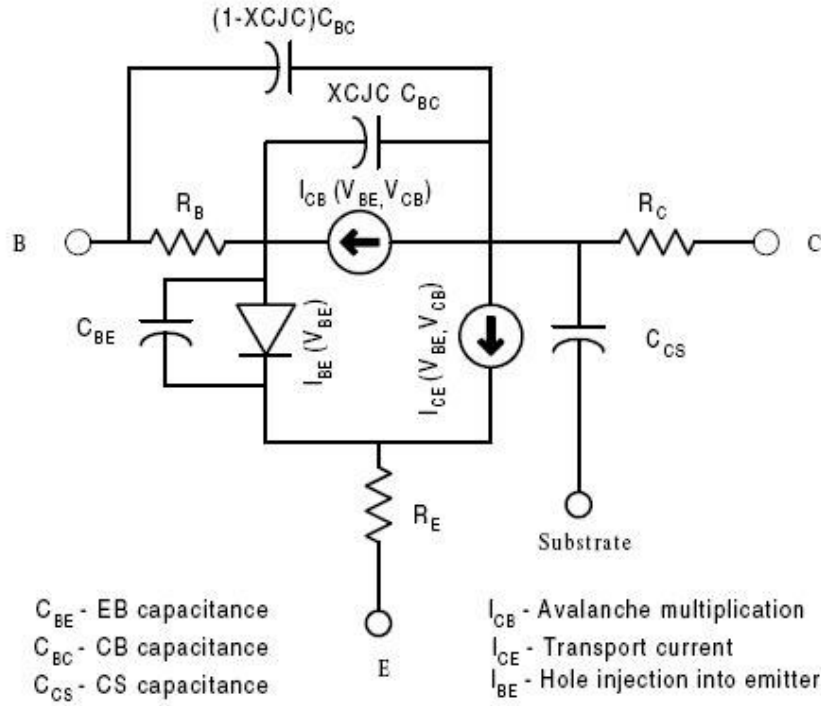
## ***2.2 Contributors of Distortion in a SiGe HBT***

Volterra Series, applied to a transistor operating as an amplifier, finds semblance to the perturbation theory employed in quantum mechanical systems. In this technique, the dominant contributors of distortion in a transistor are replaced by a linear element, in parallel with a virtual nonlinearity contributing current source. After that, solving the circuit analytically can provide us information we need to determine the overall IIP3 of the transistor/amplifier. For each distortive element, the magnitude of the above mentioned virtual nonlinearity contributing current source determines the contribution of distortion from that particular source towards overall nonlinearity. This concept becomes clearer in following passages.

Before we proceed to discuss the mathematics of VS, it will be instructive to understand the dominant nonlinearities in a SiGe HBT, as well as means to treat those analytically. In a standard forward-active region of operation for the SiGe HBT (ensuring that the transistor is still far from avalanching bias values), there exist 4 sources of nonlinearity [1], [11], [12]. They are as follows:

1. The  $I_{CE}$  exponential nonlinearity
2. The  $I_{BE}$  exponential nonlinearity
3. The  $C_{BE}$  nonlinearity
4. The  $C_{BC}$  nonlinearity

These four sources can be seen in the *npn* SiGe HBT equivalent circuit shown in Figure 6 below. Each of these will be discussed one by one in the following passages.



**Figure 6:** Equivalent circuit for *npn* SiGe HBT delineating dominant sources of nonlinearity [1].

### 2.2.1 The $I_{CE}$ Exponential Nonlinearity

This nonlinearity in a SiGe HBT stems from the exponential relationship between the base-emitter voltage  $V_{BE}$  and the output collector-emitter current  $I_{CE}$ . This results in  $I_{CE}$  being a nonlinear function of  $V_{BE}$ , resulting in a nonlinear transconductance  $g_m$  that may be described by a power series expansion of the ac current voltage relation as:

$$i_c(t) = g_m \cdot v_{be}(t) + K_{2gm} \cdot v_{be}^2(t) + K_{3gm} \cdot v_{be}^3(t) + \dots \quad (2.1)$$

Seeing the above equation, it becomes apparent that the *ac* current is the sum of a linearized current element (the  $g_m$  product) as well as second and third order nonlinear elements (denoted by  $K_{ngm}$  products). The nonlinearity coefficients are denoted by the following expression:

$$K_{ngm} = \frac{1}{n!} \frac{\partial^n f(v)}{\partial v^n} \Big|_{(v = V_{DC})} \quad (2.2)$$

Where  $V_{DC}$  is the DC component of the voltage that controls the nonlinearity function  $f$ . In this case,  $f$  is the exponential function that governs I-V relation in a transistor.

### ***2.2.2 The $I_{BE}$ Exponential Nonlinearity***

The base current  $I_{BE}$  nonlinearity is simply a scaled down version of the  $I_{CE}$  nonlinearity. The scaling factor in this case is the current gain  $\beta$ . As a result, only the nonlinearity coefficients of the nonlinear transconductance  $g_{be}$  are altered as follows:

$$g_{be} = \frac{g_m}{\beta} \quad (2.3)$$

$$K_{ngbe} = \frac{K_{ngm}}{\beta} \quad (2.4)$$



### 2.2.3 The $C_{BE}$ and $C_{BC}$ Nonlinearities

A major source of nonlinearity at high injection is the capacitive nonlinearity – dominated by base-emitter  $C_{BE}$  and base-collector  $C_{BC}$  capacitances [1]. Current is produced in an HBT due to voltage modulation of charges. This leads to formation of capacitive effects. Just as the currents described in 2.2.1 and 2.2.2, these capacitances are nonlinear functions of the terminal voltages across them. Thus the ac charge can be written in the following manner,

$$q_{ac}(t) = C \cdot v_c(t) + K_{2C} \cdot v_c^2(t) + K_{3C} \cdot v_c^3(t) + \dots \quad (2.5)$$

Here again,  $C$  is the small-signal linear capacitance. The net total capacitance however (obtained by taking a derivative of charge  $q_{ac}$  with respect to controlling voltage  $V_C$ ) is the sum of this linearized element with and second and third order nonlinear capacitances (denoted by  $K_{nC}$  terms in the equation). The nonlinearity coefficients are denoted by the following expansion:

$$K_{nC} = \frac{1}{n!} \frac{d^n f(v)}{dv^n} \Big|_{v=V_C} \quad (2.6)$$

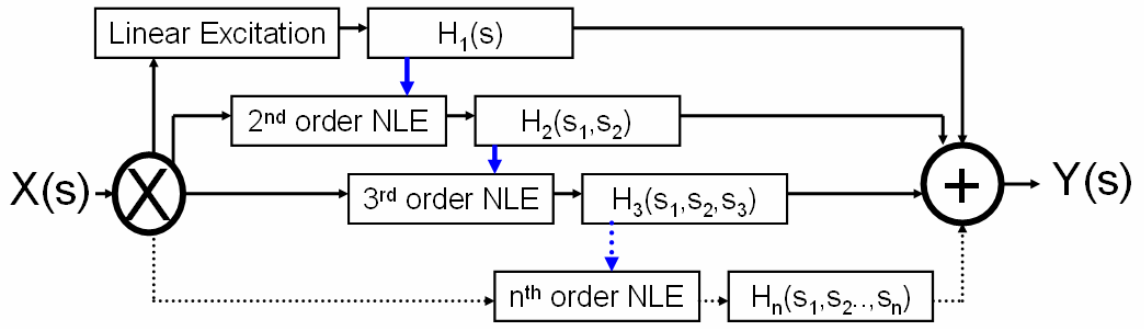
Where  $V_C$  is the controlling DC voltage applied between the two terminals (E-B or C-B) that determines the respective capacitances between those two terminals. Thus the

capacitances also become a nonlinear function of applied bias across terminals, as opposed to being constant.

### ***2.3 Volterra Series Applied to Distortion***

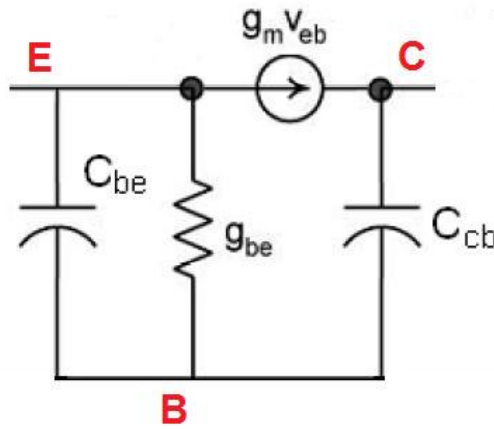
The VS approach can be utilized to solve a circuit for its  $n^{\text{th}}$  degree nonlinearity response. However, we hardly go beyond the third order nonlinearity (the IIP3), because the contribution of the higher order terms to overall nonlinearity decreases rapidly beyond the third order. The mathematics of VS has been dealt in depth in [6] - [8], and the reader is encouraged to peruse those for reference. We will go over only the bare-bones essential mathematics in this thesis.

Since Intermodulation Distortion is a frequency dependent phenomenon, the analysis is eased if carried out in the frequency (Laplace) domain. In the VS model, the response of a nonlinear system  $Y(s)$  to an input  $X(s)$  is equal to the sum of responses of a series of inputs. These inputs themselves are a combination of a linearized input, as well as virtual nonlinear current source elements of different orders. This is shown clearly in Figure 7. The input excitation  $X(s)$  can be mathematically divided into a linear excitation, as well as a host of higher order non-linear excitations (NLE). Each  $(n+1)^{\text{th}}$  order NLE has to be determined from the previous  $n^{\text{th}}$  order transfer function  $H_n(s_1, s_2, \dots, s_n)$ . The output  $Y(s)$  is the sum of product of each excitation (linear or nonlinear) with its respective transfer function. Admittedly, the mathematics of this may become tricky, but fortunately it isn't crucial to understanding transistor level distortion phenomenon.



**Figure 7:** Algorithmic flow of the Volterra Series.

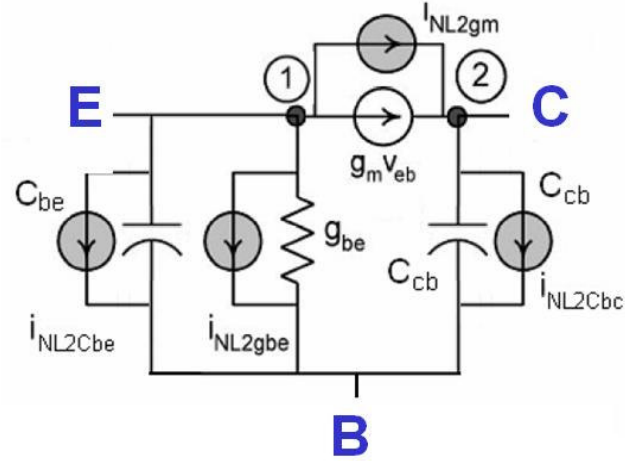
In figure 7 shown above,  $H_1(s)$  is the basic transfer function of a linearized circuit. In the case of a transistor, it is solved by exciting the circuit with a linear input voltage excitation and solving for the output using compact modified nodal analysis (CMNA) [13].



**Figure 8:** Simplified equivalent  $\pi$ -circuit of a transistor incorporating the four major nonlinearity sources.

Figure 8 shown above shows the simplest possible  $\pi$ -equivalent circuit of a SiGe HBT. The current nonlinearities are shown in their *ac* representations; hence  $I_{CE}$  nonlinearity is replaced by  $g_m \cdot v_{be}$ , while  $I_{BE}$  is replaced by  $g_{be}$ . The transistor can be biased in any of the three possible configurations; Common Emitter, Common Base, and Common Collector. Input and output nodes can thus be defined based on the configuration in which the transistor is operating. The transfer function relating the input voltages to the output voltages for a first order analysis is defined by  $H_1(s)$  in our case. CMNA can be performed on this circuit to obtain the output. The parameters ( $C_{BC}$ ,  $C_{BE}$ ,  $g_m$  etc) of the elements of this equivalent circuit have been extracted using the techniques described in [14].

Having determined the value of  $H_1(s)$ , the next step is to solve the circuit for the second order. This is done by solving the same circuit in Figure 8 but with different excitations. The excitations in this case are the second order NLEs, or the virtual non-linearity current sources. These are displayed for clarity in Figure 9 by the  $i_{NL2\_}$  terms. Each linear element in the original circuit (i.e  $g_m \cdot v_{be}$ ,  $g_{be}$ ,  $C_{BC}$  and  $C_{BE}$ ) is now added with a parallel corresponding virtual nonlinear current source.



**Figure 9:** Second order transistor equivalent circuit for use in CMNA.

It is at this point that the advantages of VS start coming to fore, in comparison to the HB approach. What the  $i_{NL2\_}$  terms achieve, in principle, is quantification of the contribution to overall circuit nonlinearity from each of these four distortion contributors. This enables us to decouple the various sources of nonlinearity acting in a transistor for a given set of  $dc$  and  $ac$  inputs. We can turn each (or all) contributor of distortion off by putting the  $i_{NL2\_}$  terms to zero, and can still compute the effect this has on the overall circuit nonlinearity. This ability to decouple nonlinearity contributors is the sole reason for VS being such a power analytical tool.

As seen in the equivalent circuit above, the circuit excitations are now the virtual nonlinear current sources. As discussed before in Figure 7, the value for these sources has to be obtained using the first order transfer function  $H_1(s)$ . Using this information, the second order voltage transfer function  $H_2(s_1, s_2)$  can easily be determined for this circuit. Again,  $H_2(s_1, s_2)$  is used to determine the third order excitations  $i_{NL3\_}$  terms. Since the

equivalent circuit for third order analysis is same as Figure 9, it is not drawn here. Having determined the excitations  $i_{NL3\_}$ , the third order voltage transfer function  $H_3(s_1, s_2, s_3)$  can easily be determined. All this mathematics is dealt in detail in reference [1].

In summary, we have been able to obtain voltage transfer functions unto the third degree for a transistor operating as an amplifier. For a 2 tone input ( $\omega_1$  &  $\omega_2$ ) this gives an analytical handle to determine the outputs at the transistor at not only at these fundamental frequencies, but also at the second order terms and the third order intermodulation terms (governed by the frequencies  $2\omega_1 - \omega_2$  and  $2\omega_2 - \omega_1$ ). All we need is the voltage or power level of the input signal, and the product of that value with the above discussed transfer functions will give us the voltage magnitude of amplified fundamental signal, as well as the 3<sup>rd</sup> order intermodulation terms at the output. Having this information at hand, finding an analytical expression for IIP3 becomes all too easy.

Putting the above discussed information in terms of analytical formulae, for a given input signal  $A \cdot (\cos\omega_1 t + \cos\omega_2 t)$ , we have:

$$V_{fundamental} = A \cdot H_1(j\omega_1) \quad (2.7)$$

$$V_{IM3} = \frac{3}{4} A^3 H_3(j\omega_1, j\omega_1, -j\omega_2) \quad (2.8)$$

Reminding ourselves that the IIP3 is a theoretical construct defined when the fundamental and the third order intermodulation terms attain equal powers, it becomes straightforward to figure out the input voltage signal at which this happens. This is done by equating equations 2.7 and 2.8 to obtain the input signal voltage  $A_{IP3}$ .

$$A_{IP3} = \sqrt{\frac{4}{3} \frac{|H_1(j\omega_1)|}{|H_3(j\omega_1, j\omega_1, -j\omega_2)|}} \quad (2.9)$$

From reference [1] and [6], the IIP3 number can be defined as follows

$$IIP3 = \frac{1}{6R_s} \cdot \frac{|H_1(j\omega_1)|}{|H_3(j\omega_1, j\omega_1, -j\omega_2)|} \quad (2.10)$$

Where  $R_s$  is the source resistance. In dBm,

$$IIP3_{dBm} = 10 \cdot \log(10^3 \cdot IIP3) \quad (2.11)$$

Thus, we have arrived at the IIP3 value for the case of a transistor operating as an amplifier. The inputs to this formula are source and load resistance, I-V and C-V characteristics, and RF excitations. The formula for IIP3 is best left in the form shown in equation 2.10 because tracing this expression back to inputs results in a formula that is multiple pages long, and offers no insight into what the nonlinearity is being caused by.

An attempt is made to simplify this expression in the next chapter, and the reader is directed to section 3.3 for that analysis.

## ***2.4 Summary***

In this chapter, the dominant contributors to distortion in a transistor, and means to treat those analytically, were discussed. The theory of the Volterra Series technique was discussed without going too much into detailed mathematics. For that, the reader is referred to relevant references.



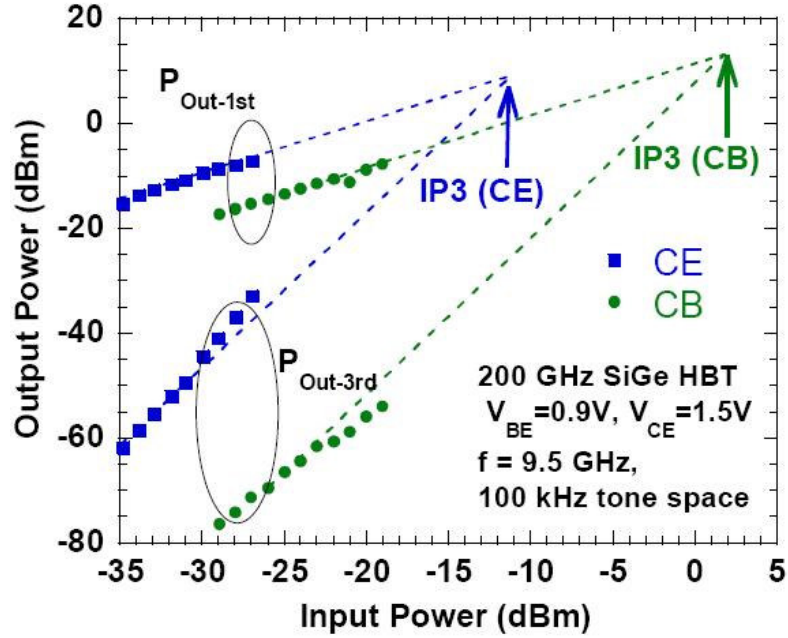
## CHAPTER 3

### DISTORTION IN COMMON-BASE CONFIGURATION

In this chapter, the linearity characteristics of a SiGe HBT operating in a common-base (CB) configuration are addressed. The CB configuration can facilitate increased collector voltage bias and dynamic voltage swing compared to common emitter (CE) topologies for applications requiring high output power [15], [16]. Additionally, it finds extensive use in other circuit topologies like cascode [17]. Hence it becomes absolutely imperative to study its linearity response both experimentally and analytically. Parts of this work were presented in 2008 *IEEE Bipolar/BiCMOS Circuits and Technology Meeting* [18], and some parts are pending journal publication in *IEEE Transactions of Microwave Theory and Techniques* [19].

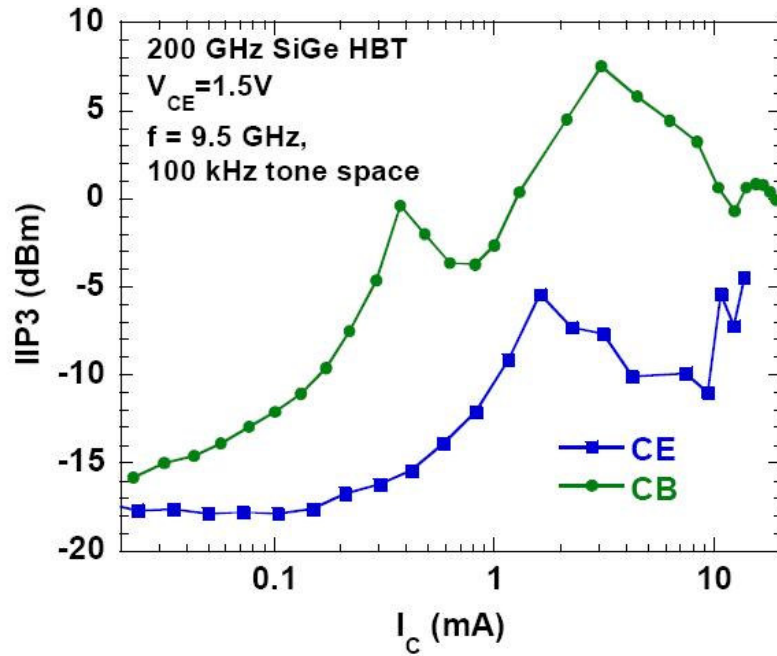
#### ***3.1 Comparison of CE and CB Linearity***

We begin by doing a comparison of linearity in CB configuration with the ubiquitous common-emitter (CE) mode of operation. The study of CE linearity has been given much attention in recent years, both the analytical aspects [20] - [22], as well as experimental data study [23], [24]. It thus becomes a priority to benchmark the linearity performance of CB in comparison to CE configuration, and that aspect has been dealt with in the next two figures. The transistors under test are IBM's 3<sup>rd</sup> generation high performance *npn* SiGe HBTs with emitter area  $A_E = 0.12 \times 3 \mu\text{m}^2$ .



**Figure 10:** The fundamental ( $P_{\text{out-1}^{\text{st}}}$ ) and 3<sup>rd</sup> order intermodulation term ( $P_{\text{out-3}^{\text{rd}}}$ ) vs.

Input power for CE and CB configurations [18].



**Figure 11:** Comparison of CE and CB linearity across bias [18].

Seen in Figure 10, a power sweep is performed on CE and CB SiGe HBTs under similar bias conditions, and the fundamental and third order intermod terms are observed

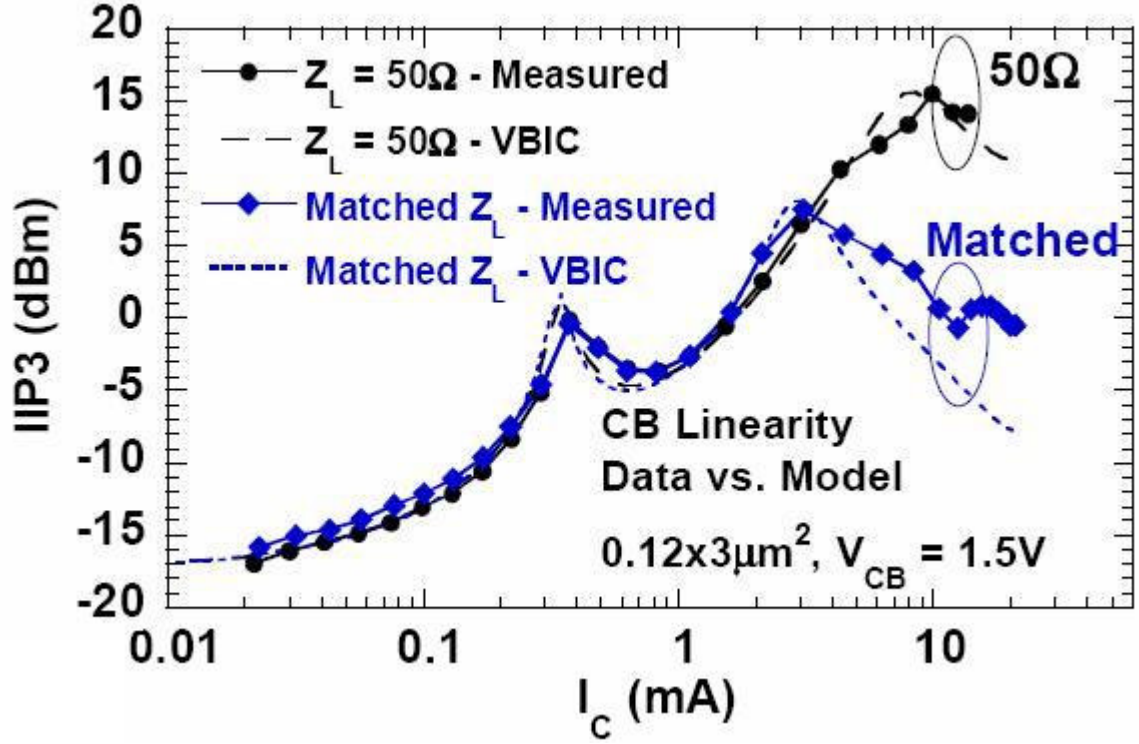
at the output. It becomes apparent that the CB configuration provides us a higher linearity than the CE (evident by larger IP3 values). In CB mode, the power gain is relatively lower than the CE case (11.4 dB compared to 20.5 dB). However, the third order intermodulation term is considerably lower in the CB configuration. This leads to substantial improvement in both IIP3 and OIP3, a fact observed in Figure 10 too.

Since the linearity of a transistor is a complex function of applied bias [1], it becomes instructive to observe the 2-tone response across bias for both topologies. This is done in Figure 11, at an input power level that is sufficiently backed off from the compression point P1dB to ensure small signal operation. An improvement in linearity is observed across bias for the CB case. This analysis is done for identically sized devices with identical *dc* and RF bias. It was concluded in [20] that the improved linearity in the CB case is due to higher input current drive (and thus higher  $P_{IN}$ ) as compared to CE configuration.

### ***3.2 Compact Model Validation for CB Linearity***

In section 3.1, we experimentally determined the differences in linearity responses of CB and CE topologies. Since the circuit designers will rely extensively on circuit simulators that utilize compact transistor models, it becomes instructive to study the fidelity of the compact model by comparing simulated linearity performance to measured data. This will not only ensure predictable circuit operation, but give us an insight into

the model's abilities to capture second-order effects like linearity. The compact model studied here to correlate with the measured linearity data is the VBIC model [25].



**Figure 12:** Measured data to model (VBIC) comparison for the CB topologies at varying load impedances [18].

From Figure 12, a number of things become clear. Firstly, the VBIC model very closely matches measured IIP3 across a broad range of bias, as well as a wide range of load impedances. “Matched” impedance was found using the Load-Pull technique [26], and is the load impedance at which the transistor displays maximum power gain. Due to very good model to measured data fidelity established by Figure 12, we can safely assume that the models will predict the linearity response from physical transistors fairly accurately for other set of stimulus conditions too.

Secondly, the recurring “sweet-spot” in IIP3 with respect to  $I_C$ , occurring at 0.4 mA, and then again at 10 mA (3 mA for “matched” case) is not an artifact of measurement, but is also captured in simulation as well. This peak can be exploited for its high IIP3 as compared to IIP3 values in the vicinity. There have not been any formulae developed to this date that can help the designer to readily calculate this “sweet-spot” bias current. This, however, is an exercise that can be made possible by employing the Volterra Series approach. In the later sections, a very simple expression for the “sweet-spot” bias current is derived and discussed.

Before arriving at compact linearity expressions, it was necessary to ensure that the Volterra Series is actually predicting the IIP3 of a transistor that was both measured experimentally and simulated using compact models. Presently, we have three methods to determine the linearity of a transistor:

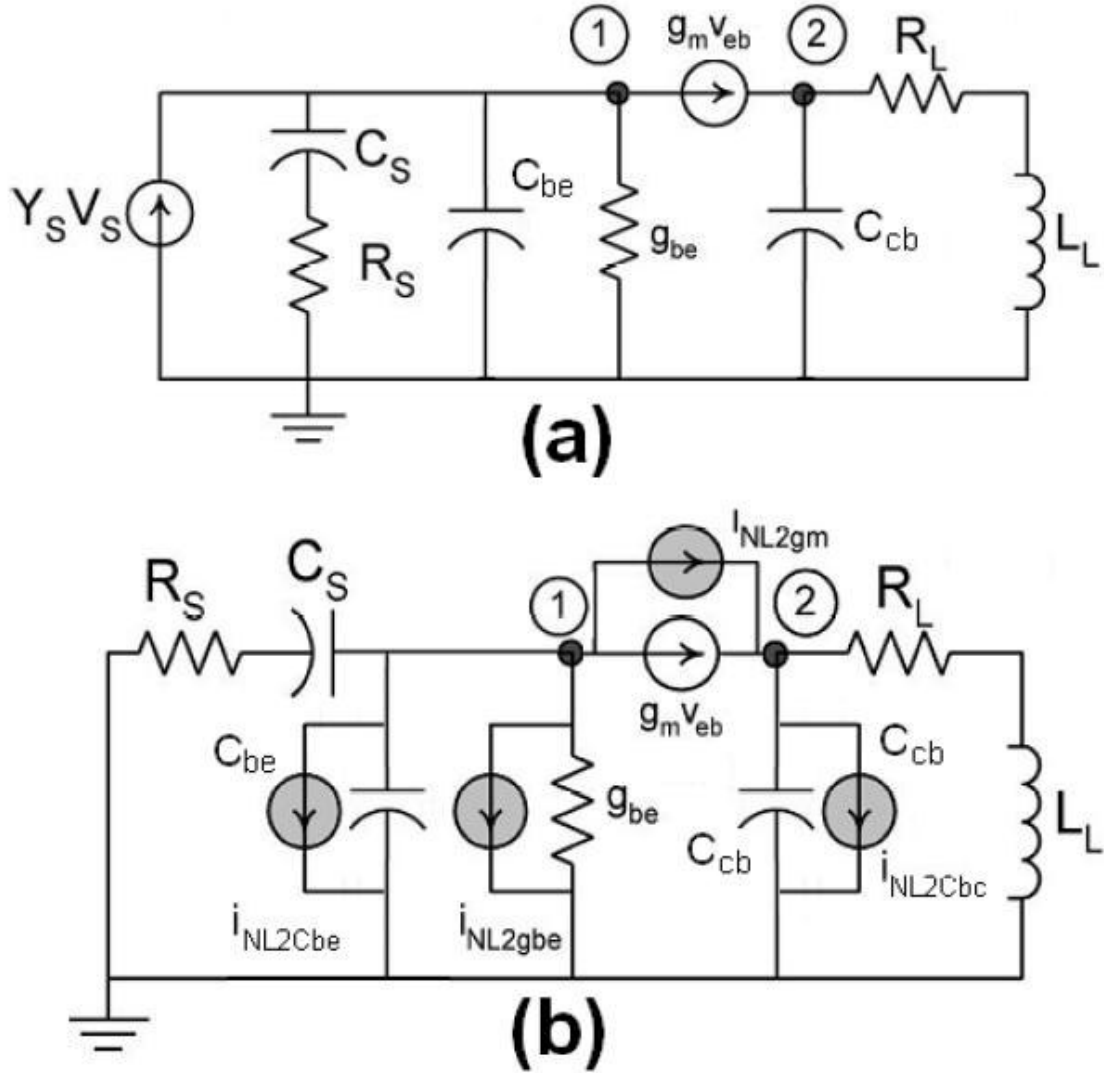
1. Experimental approach – to measure a transistor and obtain its IIP3 data
2. Compact model simulation – using commercial simulators like HP-ADS to determine transistor linearity
3. Volterra Series calculation – using the methods discussed in section 2.3 to arrive at transistor linearity response

Figure 12 established that methods 1 and 2 give us similar results. So we will rely on linearity responses from method 2 to validate calculated results obtained from using method 3 – the VS method.

### ***3.3 Calculation of CB Linearity using Volterra Series***

Using the equations described in section 2.3, an extensive program was written in MATLAB to read the input data for the transistors (I-V, load and source impedances, capacitances etc) and solve a host of equation recursively to obtain the IIP3 values. Once a match between VS and VBIC simulations was established, a set of equations was arrived at using Volterra Series that would accurately capture linearity when fed with similar inputs as above. Since these equations were bulky and spanned many pages, logical assumptions were made to simplify them. That done, a simple expression was developed that determined the “sweet-spot” of linearity (as discussed in section 3.2 and shown in Figure 12).

The equivalent circuit, developed specifically for the CB configuration, is shown in Figure 13. The CMNA equations were written up for this specific topology, and the first order circuit was solved. Using the transfer function obtained from this, second and third-order circuits were obtained to solve for the linearity values. All this was done in accordance with the methodologies discussed in section 2.3.



**Figure 13:** Simplified network used to solve the (a) first-order and (b) second-order transfer functions for a CB SiGe HBT. The third-order network is the same as (b) with “ $i_{NL2\_}$ ” nonlinear current sources replaced by “ $i_{NL3\_}$ ” sources [19].

Referring to equation 2.10 from section 2.3, IIP3 was found to be:

$$IIP3 = \frac{1}{6R_s} \cdot \frac{|H_1(j\omega_1)|}{|H_3(j\omega_1, j\omega_1, -j\omega_2)|} \quad (3.1)$$

Solving for the transfer functions for the CB case using circuit in Figure 13 gives us the expression:

$$IIP3 = \frac{|A|}{\left| 6R_S^2 (i_{NL3gm} (1 - A) - (i_{NL3gbe} + i_{NL3cbe}) A - i_{NL3cbc} \right|} \quad (3.2)$$

Where

$$A = g_m / Y_{in}(\omega_1) \quad (3.3)$$

$$Y_{in}(\omega) = 1/R_S + g_{be} + g_m + j\omega C_{be} \quad (3.4)$$

Expressions for the third-order virtual nonlinear current sources ( $i_{NL3\_}$ ) were calculated using VS and simplified using the assumption  $\omega_1 = \omega_2$ . Thus,  $g_m$  nonlinearity current source can be written as

$$i_{NL3gm} = \frac{1}{B} \left[ K_{3gm} - \frac{2}{3} K_{2gm}^2 \left( \frac{2}{Y_{in}(0)} + \frac{1}{Y_{in}(2\omega_1)} \right) \right] \quad (3.5)$$

$$B = R_S^3 Y_{in}(\omega_1)^2 Y_{in}(-\omega_1)$$

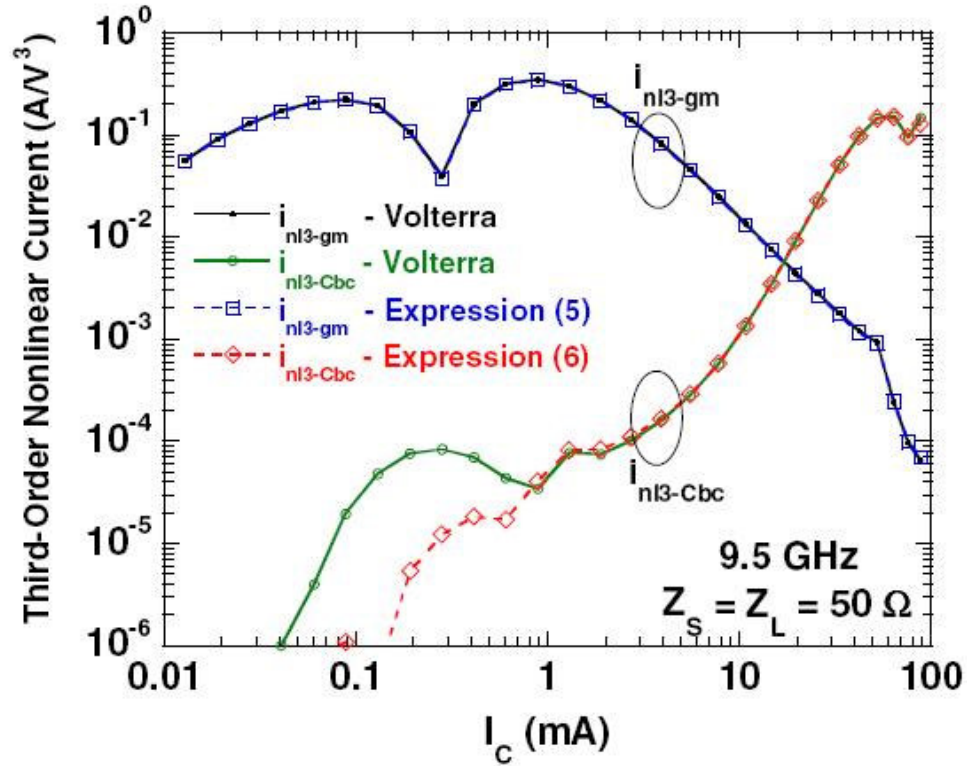
The above expression was arrived at after a lot of simplifications, and it displays that  $i_{NL3gm}$  depends on both second and third order  $g_m$  values (ref equation 3.5). Also, for an increasing input admittance  $Y_{in}$ , the nonlinearity due to exponential I-V decreases!



Similarly, after many simplifications,  $C_{BC}$  nonlinear current source can be written as:

$$i_{NL3Cbc} \approx \frac{-j\omega_1 K_{3Ccb}}{B} \frac{(g_m Z_L)^3}{(1 + j\omega_1 C_{CB} Z_L)} \quad (3.6)$$

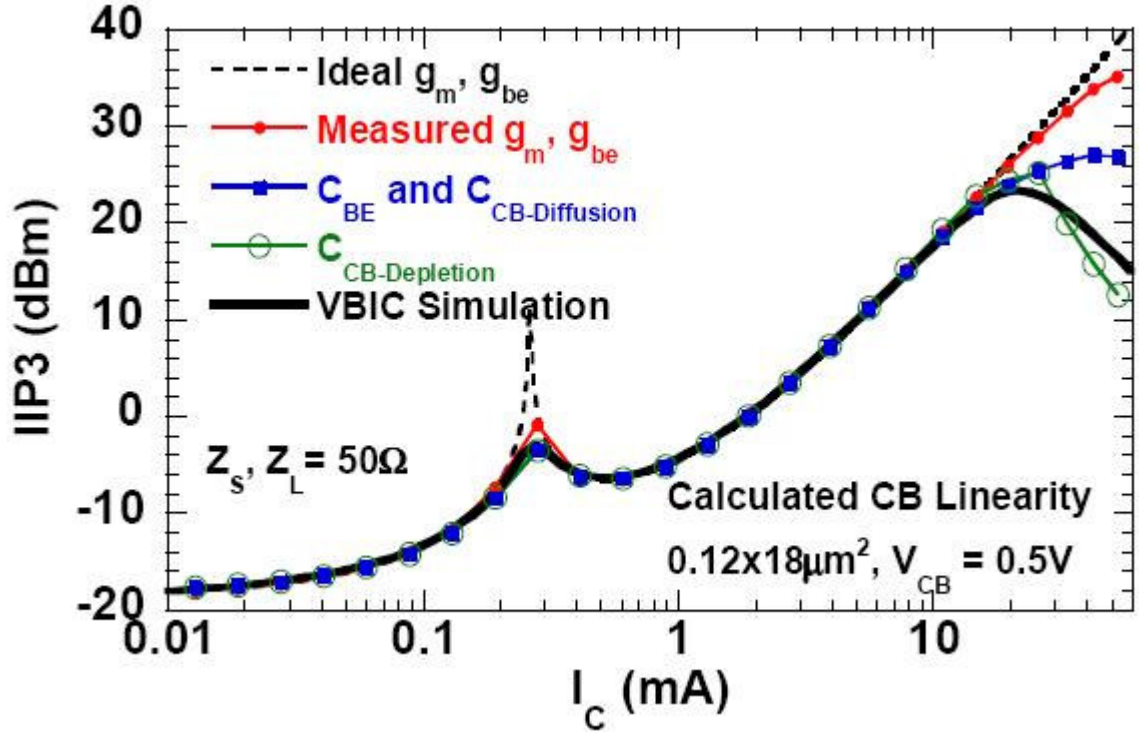
Having obtained expressions 3.5 and 3.7 after a series of assumptions, the reader might question the justification behind these assumptions, wondering to what extent they change the original bulky page-long Volterra expressions. The simplified expressions for third-order nonlinear current sources are shown in comparison with their bulky counterparts graphically in Figure 14.



**Figure 14:** 3<sup>rd</sup> order virtual nonlinear current sources for nonlinear  $g_m$  and  $C_{BC}$ , comparing the full Volterra expressions with the simplified expressions 3.5 (5) and 3.6 (6) [19].

It can be clearly seen that the simplified expression for  $i_{NL3gm}$  exactly overlays the full Volterra expression, while the simplified expression for  $i_{NL3CBC}$  overlays in the regions where we want it to, namely the high current regions where  $C_{BC}$  capacitance dominates linearity response. Hence the assumptions we made, and the expressions 3.5 and 3.6 that we derived from them, stand justified.

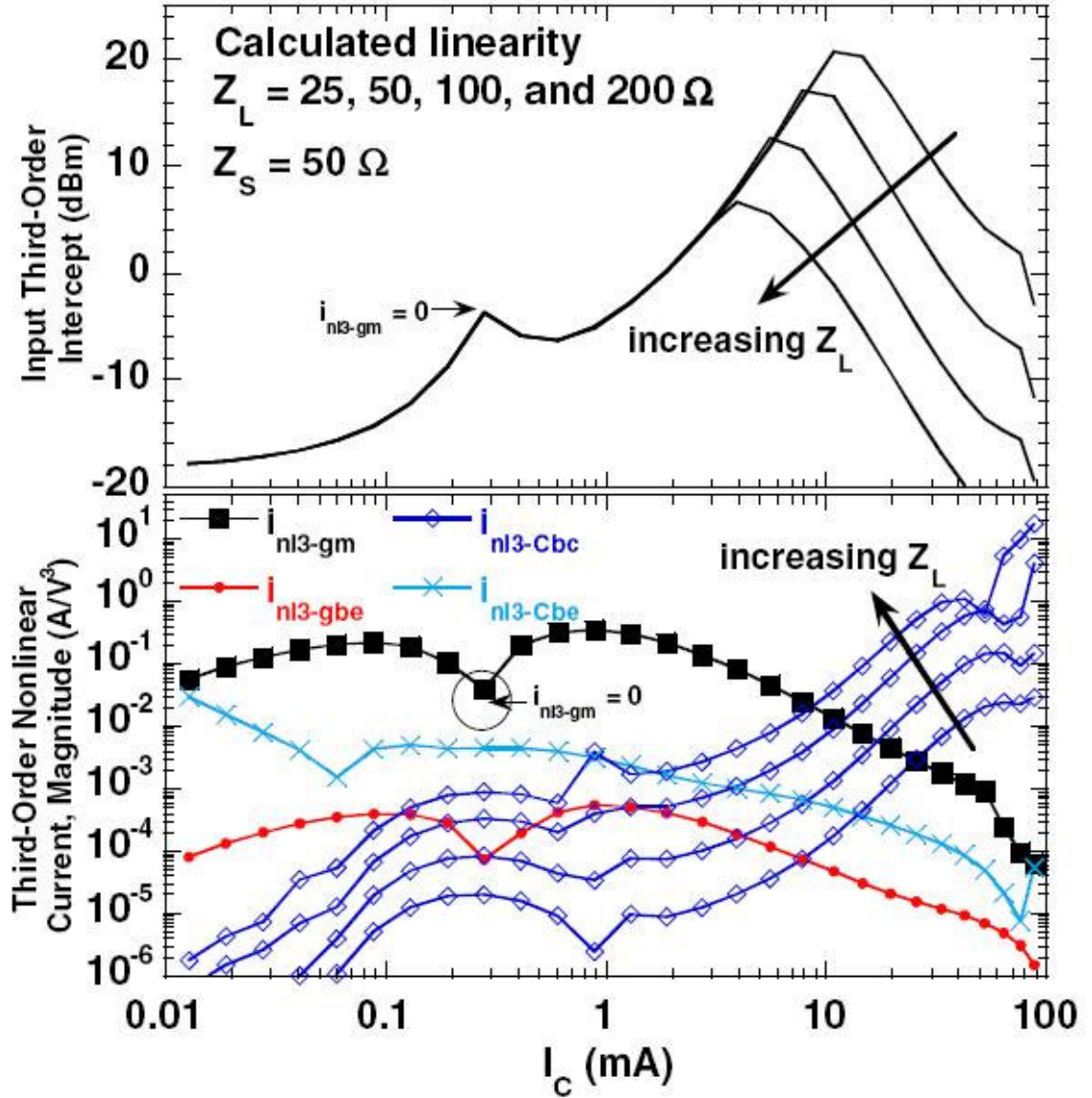
In Figure 15, shown are the linearity results obtained from simplified VS expressions developed above, in comparison with VBIC simulation results. The biggest advantage of the VS technique, the ability to decouple the distortion contributors from one another, was applied here. It is seen that with all sources turned on (green circled line), the VS method gives same linearity as VBIC simulation, again lending the calculations credibility. As we start to turn off contributors one by one, we see that linearity starts to improve in the high injection region. In the low injection region, turning off capacitive nonlinearities ( $C_{BE}$  and  $C_{BC}$ ) has no effect, as shown by the red curve. The take away from this important figure is that at low current, CB nonlinearity is dominated by exponential nonlinearity ( $g_m$ ), while at high currents, it is dominated by capacitive ( $C_{BC}$ ) nonlinearities.



**Figure 15:** CB linearity calculated using the VS approach, and compared with simulation from VBIC model [18].

### 3.4 Understanding Role of Load Impedance on Distortion

In Figure 12, we saw from measured data that the IIP3 of a transistor changes dramatically with applied bias depending on the load impedance. It is attempted here to understand that phenomenon analytically using the VS model developed in previous section.



**Figure 16:** Calculated CB linearity performance over bias for a SiGe HBT at different  $Z_L$ . Top curve shows only IIP3 vs.  $I_C$  with varying  $Z_L$ . Bottom curve shows the virtual nonlinearity current sources with changing  $Z_L$ . Only  $i_{NL3Cbc}$  changes with  $Z_L$  [19].

As observed in Figure 12, IIP3 reduces with increasing  $Z_L$  but only in the high current region. This is seen in Figure 16 as well. The bottom graph shows how the virtual nonlinear current sources due to the four possible nonlinearity contributors are acting

with changing bias. Because  $i_{NL3gm}$  has the highest value in the low bias region (till 10 mA), exponential nonlinearity ( $g_m$ ) is the dominant nonlinearity here. Beyond 10 mA,  $i_{NL3Cbc}$  has the highest value and hence becomes the dominant contributor to overall nonlinearity. Only this value changes with the load impedance  $Z_L$ , also corroborated by equation 3.6. The takeaway here is simple: for increasing  $Z_L$ , the  $C_{BC}$  capacitance becomes more and more nonlinear in the high current region (marked by increased  $i_{NL3Cbc}$ ), which leads to diminishing IIP3 in the high current region with increasing  $Z_L$ .

In the low current region, the oft talked about linearity “sweet-spot” is seen. The reader can see that this bias current coincides with the point where  $i_{NL3gm}$  has a zero value in the bottom graph. This sweet spot can be utilized in circuit design for its high linearity, and a simplified expression to get at the sweet-spot is discussed below.

From equation 3.5,  $i_{NL3gm}$  can go to zero (neglecting frequency effects of  $C_{BE}$ ) when:

$$K_{3gm} = 2K_{2gm}^2 / Y_{in}(0)$$

Assuming  $g_m = I_C/V_T$ , the peak in IIP3 occurs when

$$I_{C(i_{NL3gm}=0)} = \frac{V_T}{R_S(2 - 1/\beta)} \quad (3.7)$$

Where  $R_S$  is the source resistance and  $\beta$  is the current gain of the transistor. Thus, a very compact expression for the  $dc$  bias current was developed at which the transistor would display an IIP3 peak. This information can be useful for the circuit designer to readily compute the bias value for high linearity.

### ***3.5 Summary***

In this chapter, the linearity characteristics of the CB configuration were dealt with in detail. An analytical technique using the VS approach was developed that matched very well with both measured linearity data and simulations from VBIC model. Using this technique, it was possible to extract expressions for individual nonlinearity contributors within a transistor specifically for the CB configuration. These expressions were simplified, and were used to lend insight into the linearity response of a SiGe HBT with changing load impedance  $Z_L$ . Also, a compact expression for the bias current at which the linearity “sweet-spot” will occur was developed. This can be used to design high linearity circuits.

## CHAPTER 4

### COMPARING LINEARITY OF COMPLEMENTARY SIGE HBT

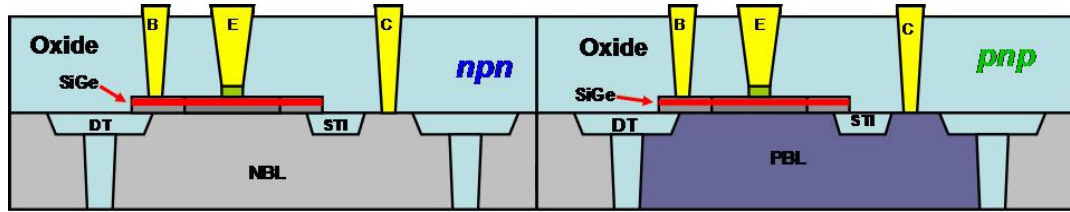
Till now, the linearity characteristics of only *nnp* SiGe HBTs have been discussed. However, a discussion on transistor linearity will not be complete until linearity response of the *pnnp* SiGe HBTs is discussed too. Using *pnnp* SiGe HBTs for high performance circuit design has recently been enabled due to availability of complementary technology platforms that provide matched electrical characteristics (similar *dc* and *ac* performances for both *nnp* and *pnnp* SiGe HBT). Hence, a comparative linearity study between *nnp* and *pnnp* SiGe HBT that empowers the circuit designers to choose better circuit building blocks is absolutely imperative. The results of such a study are shown in this chapter. This work was presented in 2009 *IEEE Bipolar/BiCMOS Circuits and Technology Meeting* [27].

#### ***4.1 Complementary Technology Overview***

In recent times, the realm of SiGe technology has seen development of high performance complimentary (*nnp + pnnp*) transistors. This has enabled a host of complementary circuit design techniques that weren't possible earlier due to inadequate electrical characteristics of the *pnnp* SiGe HBT. The development of high performance of *pnnp* transistors, whose *dc* and *ac* performance matches that of an *nnp*, has proven to be a boon for analog and mixed-signal design. Complementary circuit technique possess a number of intrinsic advantages over their *nnp*-only counterparts, namely, efficient utilization of rail-to-rail

voltages, reduced die area, reduced static power dissipation, all of which translate into more efficient circuits. A direct circuit candidate that benefits from this is the high-speed complementary push pull output stage [28], [29].

Linearity study of transistors in this chapter was done using an experimental high-performance platform developed by National Semiconductor. The full performance specs of the technology will be revealed in the near future. The technology has a full BiCMOS platform that supports *nnp* and *pnnp* SiGe HBTs with matched electrical characteristics. Achieving this has been an issue historically, due to slow *pnnp*. Both HBTs have a  $BV_{CEO} > 3.6$  V and a peak  $f_T > 30$  GHz, thus enabling utilization in high performance circuits. A rudimentary cross section of the *pnnp* and *nnp* SiGe HBTs is shown in Figure 17. It is clear that this platform utilized Deep Trench and Shallow Trench Isolation techniques.

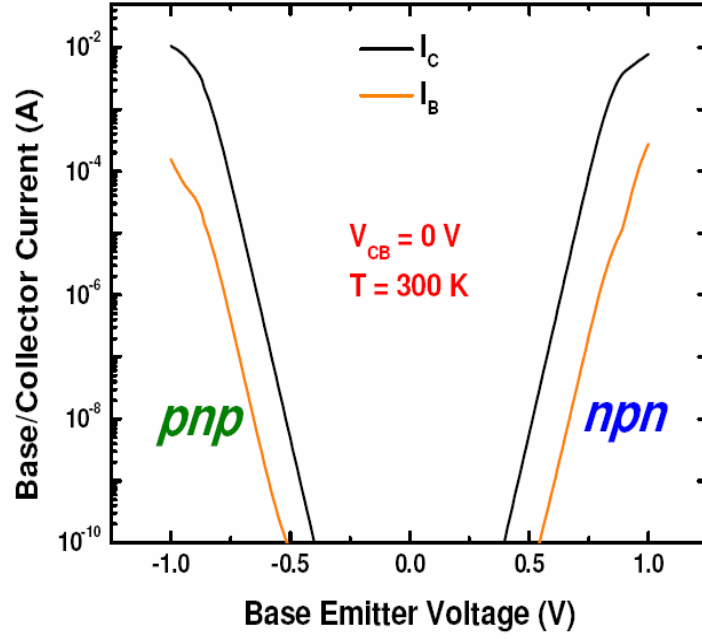


**Figure 17:** Cross-section of *nnp* and *pnnp* SiGe HBTs developed by NSC used in this study.

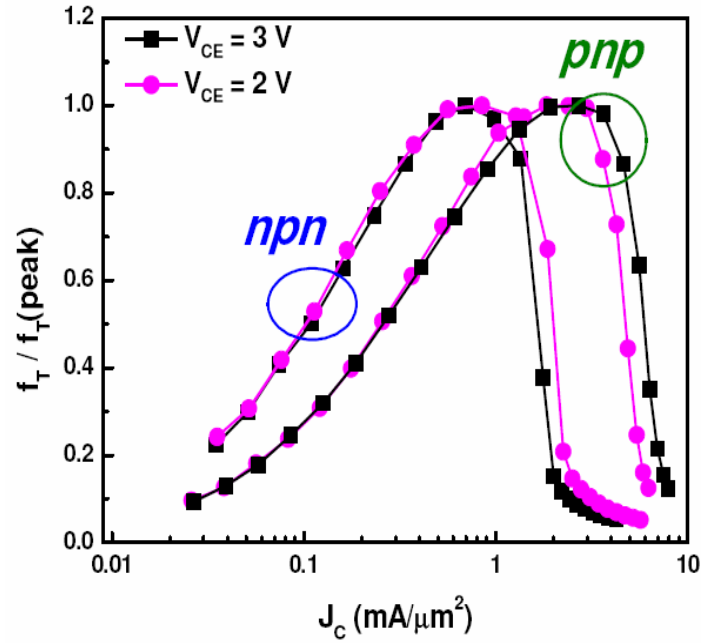
To reiterate matched electrical performance for the complementary devices, plots of *dc* and *ac* data are shown below. Figure 18 shows the gummel plots for both transistor types with similar geometries and applied bias. It is seen they display identical current



and current gain values. Figure 19 shows normalized  $ac$  performance of both transistor types with similar geometries. Again, ideal  $f_T$  responses can be observed.



**Figure 18:** Gummel plots for similar sized  $nnp$  and  $pnp$  SiGe HBTs.

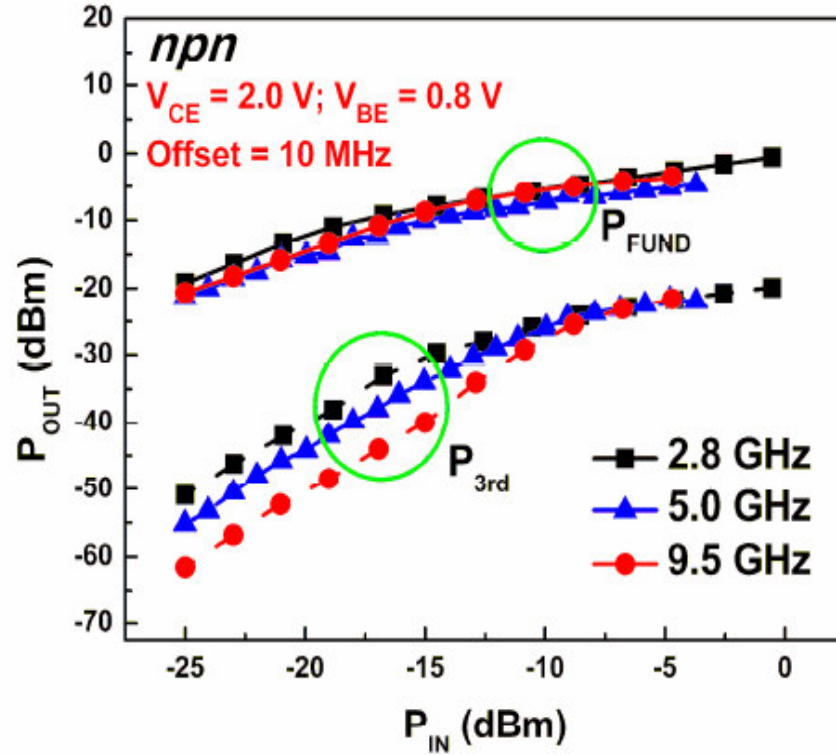


**Figure 19:** Cut-off frequencies  $f_T$  normalized to peak  $f_T$  of transistors vs current density

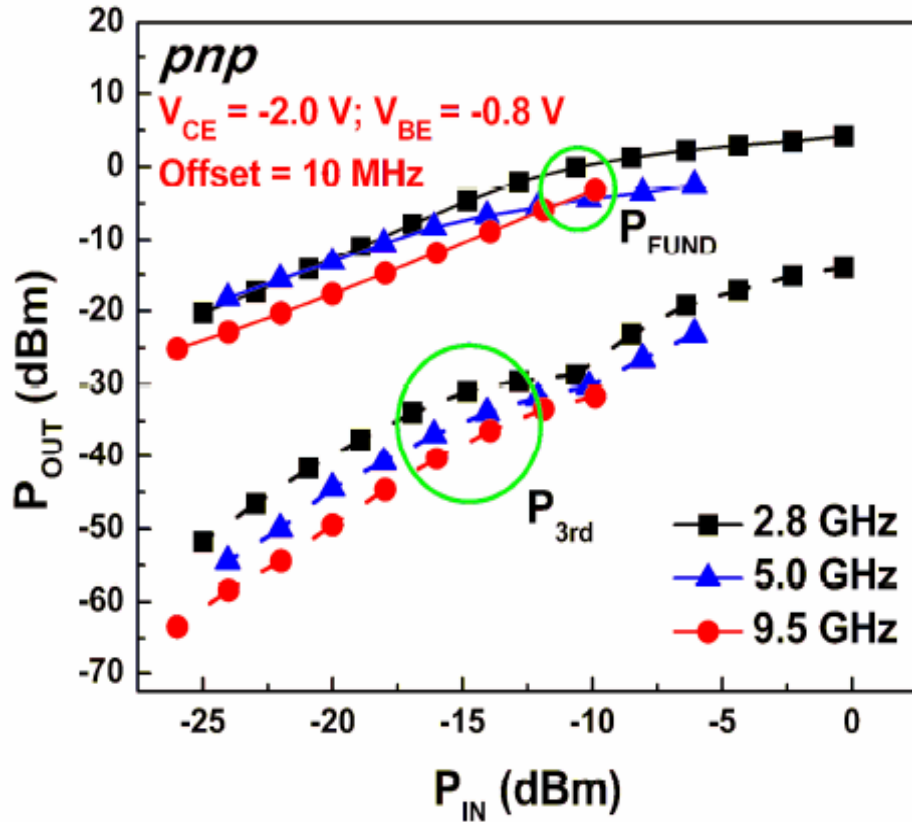
$J_C$ .

## 4.2 Two-Tone Power Sweep

Two-tone tests were done on similar sized complementary devices under similar applied bias with 50  $\Omega$  source and load terminations, for a varying number of input tones. This experiment was done to determine regions where an ideal 1:3 slope is observed between the fundamental and third order intermod terms. The transistors in this study are biased in a Common Emitter (CE) configuration.



**Figure 20:** Fundamental ( $P_{FUND}$ ) and third order intermod term ( $P_{3rd}$ ) as a function of input power  $P_{IN}$  for npn SiGe HBT.

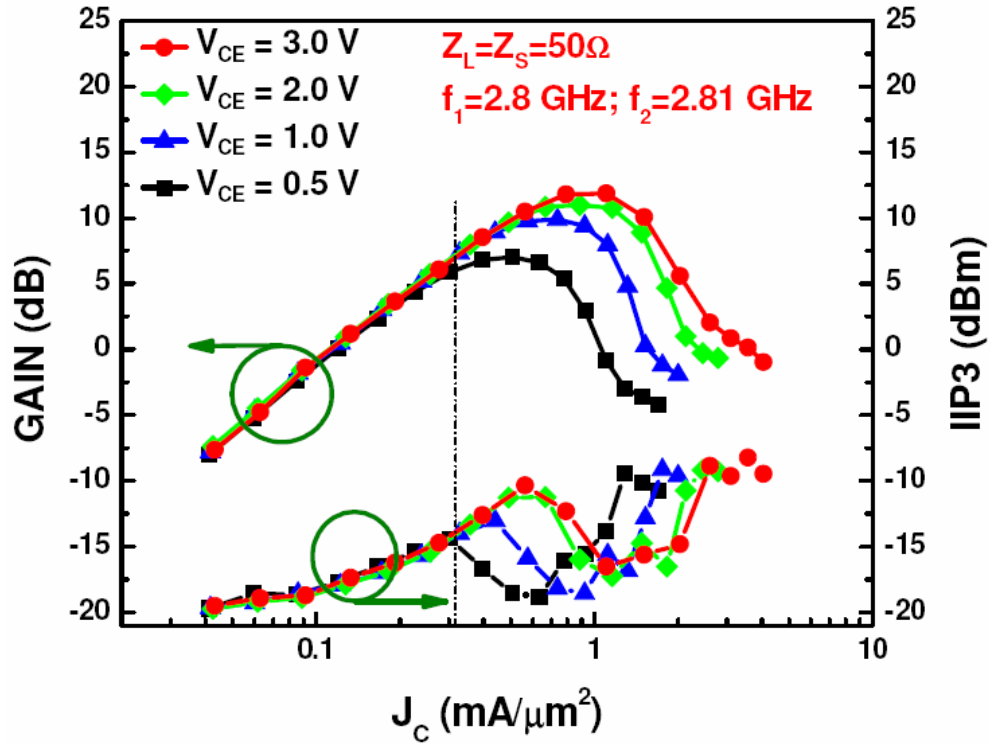


**Figure 21:** Fundamental ( $P_{FUND}$ ) and third order intermod term ( $P_{3rd}$ ) as a function of input power  $P_{IN}$  for *pnp* SiGe HBT.

An input signal level of -25 dBm is sufficiently backed off from any kind of saturation phenomenon. Hence an input tonal power of -25 dBm is utilized for further experiments in this chapter. Also, it can be seen that the 3<sup>rd</sup> order intermod terms decrease in power with increasing tone frequency. This would translate directly into increasing IIP3 with increasing input frequency. This phenomenon is discussed in [20] as well.

### 4.3 Effect of Collector Bias on Distortion

In a comparative context, the linearity response of similar sized complementary transistors with varying collector bias is studied below. The power gain and IIP3 data vs. bias current density  $J_C$  gathered for the *npn* SiGe HBT is shown in Figure 22.



**Figure 22:** IIP3 and Gain vs. current density  $J_C$  with increasing  $V_{CE}$  for standard sized *npn* SiGe HBT.

A number of things become clear from the above figure. Firstly, the gain and the IIP3 of a transistor are independent of  $V_{CE}$  in the low  $J_C$  region (region to the left of the demarcating dashed line). In the high  $J_C$  region, both the gain and IIP3 increase with an increasing  $V_{CE}$ . This is a very interesting trend, and can be explained by device physics.

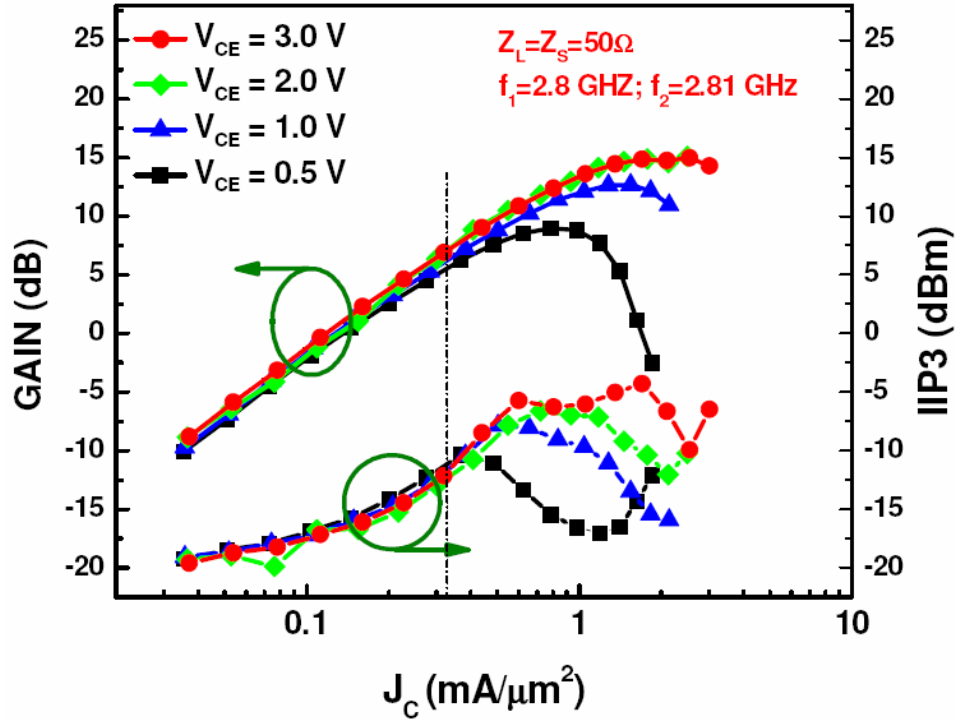
Looking at just the gain metrics for now, it can be proven that the gain is rolling off at a later bias current due to delayed onset of Kirk Effect [30] with an increasing  $V_{CE}$ . According to the classical equation in [30]

$$J_K = qv_{SAT}N_C \left[ 1 + \frac{2\epsilon(V_{CB} + \phi_{BI})}{qN_C W_{epi}^2} \right] \quad (4.1)$$

$J_K$  is the collector current where the gain of the transistor starts to roll-off. Here,  $v_{SAT}$  is the carrier saturation velocity,  $N_C$  is the collector doping density,  $V_{CB}$  is the reverse bias across the CB junction,  $\phi_{BI}$  is the built-in junction potential, and  $W_{epi}$  is the width of collector epi layer. Now, for an increasing  $V_{CE}$ , an increasing reverse bias  $V_{CB}$  is applied across the Collector-Base CB junction. With an increasing  $V_{CB}$ , the value of  $J_K$  (the gain roll-off current point) is being increased. This leads to increased gains at higher  $J_K$  before we start observing the gain degradation. This same effect is observed in Figure 22 above.

Explaining the IIP3 trends, a large reverse bias  $V_{CB}$  due to high  $V_{CE}$  will result in improved linearity performance, since the collector of the transistor is more depleted [31] - [34]. When the collector epi-layer is fully-depleted, the width of the depletion region becomes a constant equaling  $W_{epi}$  (distance from CB junction to the subcollector). This, in turn, causes the capacitance of the reverse-biased CB junction ( $C_{BC}$ ) to become independent of any change in voltage applied across its terminals. We established in chapter 3 that the rate of change of  $C_{BC}$  with respect to changing  $V_{CB}$  is the dominant source of nonlinearity in the high current region [1], this nonlinearity contributor diminishes for increasingly depleted collector epi-layers. Hence, the peak IIP3 values

improve with increasing  $V_{CE}$ , since the epi-layer is depleted more and  $C_{BC}$  is thus more resistant to changes in  $V_{CB}$ . This is consistent with the measurements shown in Figure 22.



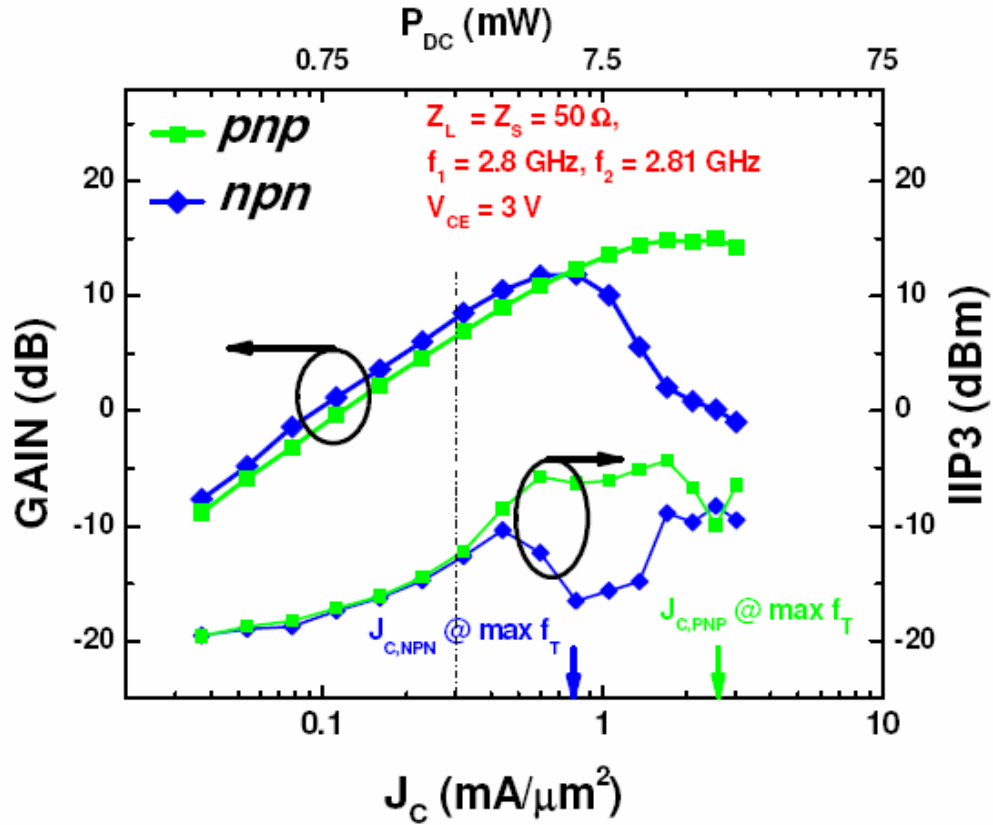
**Figure 23:** IIP3 and Gain vs. current density  $J_C$  with increasing  $V_{CE}$  for standard sized *pnp* SiGe HBT.

In Figure 23, we observe similar trends in gain and IIP3 for the *pnp* SiGe HBT too. A similar reasoning as discussed previously applies here as well.

#### 4.4 Comparison of *nnp* and *pnp* Distortion

Having determined what changing  $V_{CE}$  can do to gain and IIP3 of *nnp* and *pnp* SiGe HBTs separately, it becomes instructive to see their respective performance head-to-head, i.e. fixing transistor size and biases for both *nnp* and *pnp*, and performing a similar

analysis as above. The results of this study are shown in Figure 24 below. The load and source terminations for both transistor types were fixed at  $50\ \Omega$  to enable apples to apples comparison.



**Figure 24:** IIP3 and gain vs. current density ( $J_C$ ) with fixed bias for a standard sized *nnp* and *pnp* SiGe HBT.

Shown above in Figure 24 are gain and IIP3 of similar sized complementary transistors. To highlight the role of Kirk effect onset in the gain roll-off, current density values for maximum  $f_T$  at the same bias are marked for both *nnp* and *pnp* SiGe HBTs. It is indeed observed that gain roll off takes place at the marked  $J_C$  value. It is also seen that at low  $J_C$ , gain and IIP3 for both transistor types is identical. At high  $J_C$  region, *pnp* have

a higher gain and linearity than the *npn* SiGe HBT. This again can find explanation in how the transistors are fabricated.

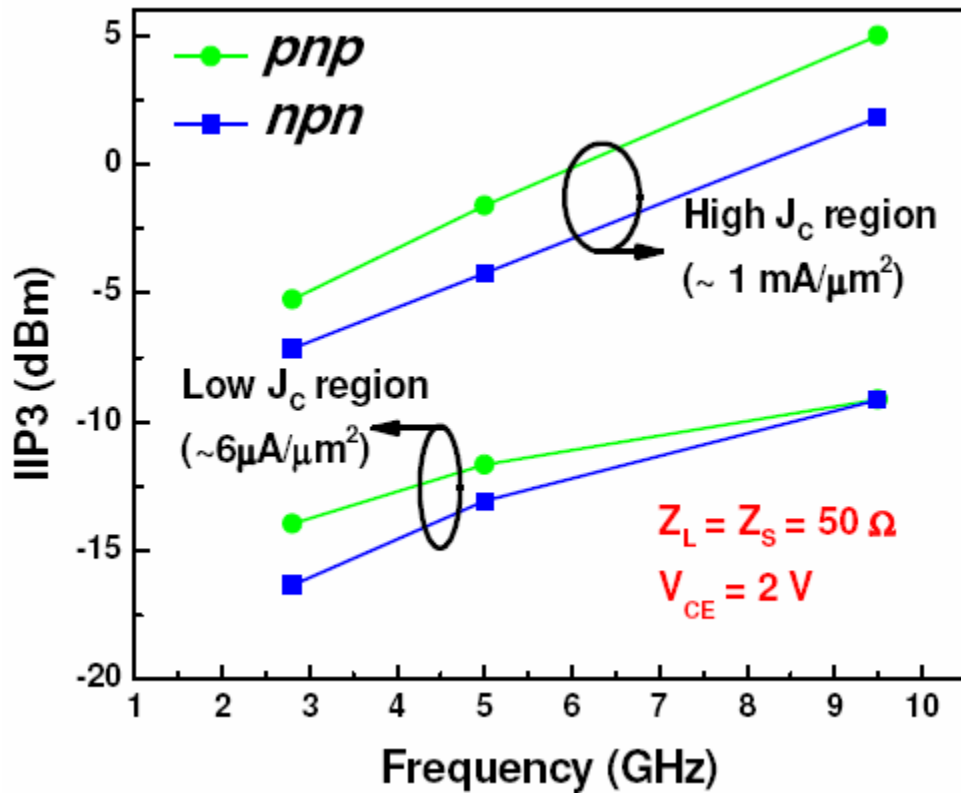
For a *pn*p SiGe HBT fabricated identically as a *n*p*n* SiGe HBT (with identical doping profiles etc), the *n*p*n* will inevitably be faster due to faster electron mobility and reduced effective mass of electrons compared to holes. Thus, to fabricate transistors with matched electrical performances (similar  $f_T$  and BV), the collector of the *pn*p will have to be doped higher than the *n*p*n*. This can be achieved without compromising the breakdown voltage of the transistors because the majority carriers in the *pn*p – holes – have lower impact ionization rates [27].

Doping the collector higher has a direct impact on the linearity and gain performance of the *pn*p SiGe HBTs. Looking at equation 4.1, a higher collector doping (higher  $N_C$ ) leads to a higher  $J_K$  value for a *pn*p transistor. Thus the *pn*p can reach higher  $J_C$  values, getting higher gains than *n*p*n* SiGe HBT, before their gain starts to drop (observed too in Figure 24). Secondly, a higher  $N_C$  makes the collector of the *pn*p more resilient to changes in applied  $V_{CB}$ . In summary, trying to make the *pn*p electrically matched to *n*p*n* naturally buys us more linearity and gain at high  $J_C$  region. Seen another way, for equal dc power  $P_{DC}$  dissipated, the *pn*p transistor will fetch us higher gain and IIP3 than the *n*p*n* device. Doping the collector higher is hence a sure-fire method to enhance linearity of a transistor. Playing with germanium profile design to enhance linearity has also been attempted, and the interested reader is encouraged to peruse [35].



#### 4.5 Effect of Scaling Frequency on Linearity

As established previously, the *pn*p and the *np*n SiGe HBTs have similar gain and IIP3 in the low  $J_C$  region, and *pn*p has higher gain and IIP3 in the high  $J_C$  region. Recognizing that these two bias regions merit separate analysis, IIP3 was observed as a function on changing input RF excitations for both types of transistors. The results of this are shown in Figure 25.

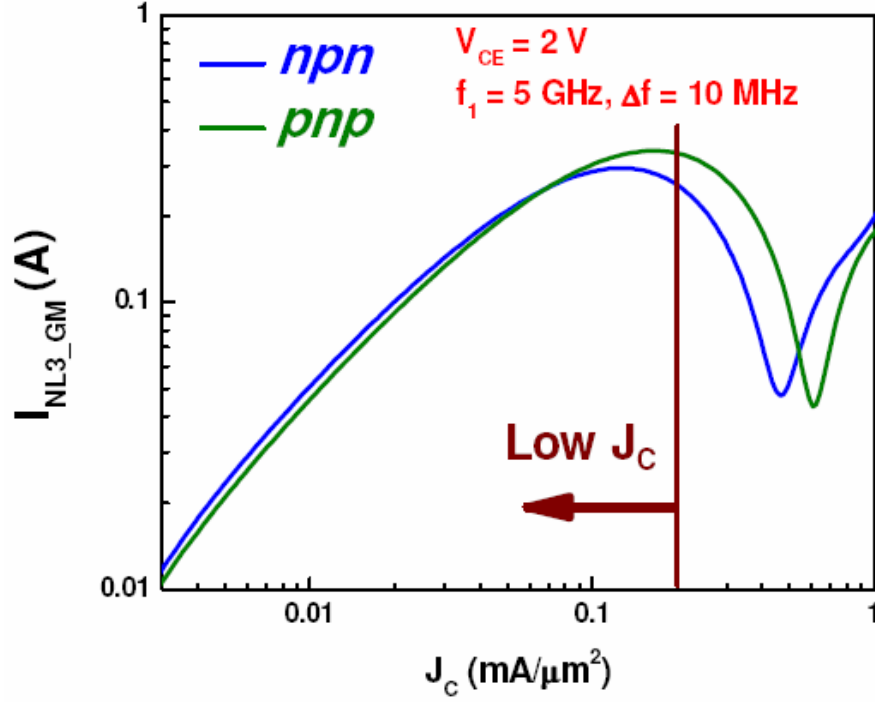


**Figure 25:** Effect of scaling input frequency on IIP3 for *npn* and *pn*p SiGe HBTs in the high and low current density region.

As expected for the low current density region, the linearity of both transistor types is very similar under similarly applied bias. For high  $J_C$  region, the linearity of the *pnp* is higher than that of the *nnp*. The explanation for this has been provided in previous sections. It might be counter-intuitive to the reader to see the linearity increasing with increasing frequency. The reason for its occurrence is that, even though the gain decreases with frequency due to increased parasitics in the transistor, the third order intermodulation terms also decrease. This is observed in Figures 20 and 21 too. Hence, the ratio of the fundamental to the third order intermod terms, which defines the overall IIP3, increases and hence the IIP3 is seen to increase with increasing input frequencies.

#### ***4.6 Understanding Complementary Linearity Response***

In this section, the explanation for similar IIP3 performance of the complementary devices is presented. From chapter 3 it was established that in the low  $J_C$  region, the dominant contributor of nonlinearity is the exponential ( $g_m$ ) nonlinearity. Thus, looking at the virtual third order nonlinear current source  $i_{NL3gm}$  for both transistor types will shed some insight into linearity in the low  $J_C$  region. This is done in Figure 26 below.



**Figure 26:**  $i_{NL3gm}$  for similar sized *nnp* and *pnnp* SiGe HBT at similar bias.

Seen above, for the low  $J_C$  region, both transistors have similar values of virtual nonlinearity current sources  $i_{NL3gm}$ . Thus, it goes on to show that linearity in the low  $J_C$  region is independent of the collector design. The collector design, however, plays a crucial role in the linearity response in the high  $J_C$  region, as discussed previously.

#### 4.7 Summary

In this chapter, the results of a comparative study of linearity for electrically matched *nnp* and *pnnp* SiGe HBTs were discussed. Linearity changes due to changing collector bias were discussed, and explanations based in device physics were furnished. It was determined how increased linearity can be obtained from a transistor. In a comparative

context, the linearity and gain of a *pn*p SiGe HBT was either equal to, or greater than, the *np*n SiGe HBT. Reasons for this based in device physics were furnished. IIP3 across frequency was examined and was found to increase with input frequency for both device types. Analytical reasons based on Volterra Series corroborated the same fact.

## **CHAPTER 5**

### **CONCLUSION**

#### ***5.1 Conclusions***

In this thesis, the distortion/linearity performance of SiGe HBTs has been given in depth attention. A three-step approach, involving simulation, measurement and analytical treatment, was applied to the problem of nonlinearity in transistors and circuits. The results discussed here will enable the circuit designers to optimize their circuits for enhanced RF linearity, thus enhancing the already high performances that SiGe technology buys for circuit design.

In chapter 1, we discussed an overview of the SiGe BiCMOS technology. We also defined the distortion problem in telecommunication systems, and discussed briefly the type of nonlinearities that affect a communication system.

In chapter 2, we discussed a mathematical tool called Volterra Series. We discussed how this powerful tool can be applied to understand the problem of distortion in a transistor, and thus find means to mitigate it.

In chapter 3, we applied the Volterra Series to understand the distortion performance of a SiGe HBT biased in a Common Base (CB) configuration. Dominant sources of nonlinearities were identified using the Volterra Series, and a very simple analytical expression was derived for the optimum-linearity bias point. This will enable the circuit designers to bias circuits for high linearity.

In chapter 4, the linearity performance of complementary SiGe HBTs was discussed. Differences in linearity performances were observed, and explanations based in device

physics were discussed. This study will give designers an enhanced arsenal of tools at their disposal to design highly linear circuits.

In conclusion, a very in-depth treatment was applied to the problem of distortion in SiGe HBTs, and methods were discussed both at device and circuit levels to mitigate the distortion problem.

## ***5.2 Future Directions***

This thesis covers the topic of distortion in SiGe HBTs for use in modern day telecommunication systems. SiGe HBTs, with continuous process innovation, have attained aggressive metrics of operation, with  $f_T$  and  $f_{MAX}$  as high as 500 GHz, thus easily rivaling their III-V counterparts while still retaining the cost advantage of CMOS platforms.

The advantages of SiGe HBTs however, are manifold, and recently SiGe HBTs have begun to be employed in Space-based systems due to their impressive performances in extreme environments like intense radiation and very low temperatures [36]. Thus, a logical extension of this work would be to study the distortion performance of SiGe HBTs under a number of extreme environment conditions – under radiation from high and low energy particles, as well as under cryogenic temperatures. This will ease the understanding of how extreme environments affect the distortion of SiGe HBTs, hence enabling their use for communication and telemetry purposes on lunar and other space-based missions.

In this thesis, an in-depth analysis of the linearity performance of a CB topology was performed using Volterra Series, and very easy expressions were derived to enable high linearity circuits. Even though the mathematical formulations of the Volterra Series are

tough, it is possible to apply the same technique to the Common Collector topology, and from there, transitioning onto distortion analysis of circuits with multiple transistors. The mathematics of the Volterra Series at some point will become too cumbersome to analyze a large number of transistors, but it will be instructive to establish the point beyond which applying VS does more harm than good.

## REFERENCES

- [1] J. D. CRESSLER, *et al.*, Silicon Germanium Heterojunction Bipolar Transistors. Boston: Artech House, 2003.
- [2] J. D. CRESSLER, Editor, The Silicon Heterostructure Handbook, CRC Press, 2006.
- [3] E. Laskin, *et al.*, "A 140 GHz double-sideband transceiver with amplitude and frequency modulation operating over a few meters," *IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 178 - 181, October 2009.
- [4] A. Joseph, *et al.*, "A 0.18 $\mu$ m BiCMOS technology featuring 120/100 GHz (fT/fMAX) HBT and ASIC compatible CMOS using copper interconnect," *IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 143-146, 2001.
- [5] L. C. N. de Vreede, "Linearization techniques at the device and circuit level," *IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 1-8, October 2006.
- [6] S. Narayanan, *et al.*, "An analysis of distortion in bipolar transistors using integral charge control model and volterra series," *IEEE Transactions on Circuit Theory*, vol. 20, no. 4, pp 341 - 345, July 1973.
- [7] D. D. Weiner, *et al.*, Sinusoidal Analysis and Modeling of Weakly Nonlinear Circuits, Van Nostrand Reinhold, 1980.
- [8] P. Wambacq, *et al.*, Distortion Analysis of Analog Integrated Circuits, Kluwer Academic, New York, 1998.
- [9] K. Kundert, *et al.*, Steady-State Methods for Simulating Analog and Microwave Circuits, Norwell, MA: Kluwer Academic Publishers, 1990.
- [10] S. Maas, "What you need to know about Volterra-Series analysis," MW Office Application Notes, 1999.
- [11] L. C. N. de Vreede, *et al.*, "Optimum dimensions of the epilayer for third-order intermodulation distortion," *IEEE Bipolar/BiCMOS Circuits and Technology Meeting 1998*, pp. 168-171, September 1998.



- [12] L. C. N. de Vreede, *et al.*, "Bipolar transistor epilayer design using the MAIDS Mixed-Level simulator," *IEEE Journal of Solid State Circuits*, vol. 34, no. 9, pp. 1331-1338, September 1999.
- [13] G. Gielen, *et al.*, Symbolic Analysis for Automated Design of Analog Integrated Circuits, Kluwer, New York, 1991.
- [14] P. Baureis, *et al.*, "Parameter extraction for HBT's temperature dependent large signal equivalent circuit model," *Tech. Dig. IEEE GaAs IC Symposium*, pp. 263-266, 1993.
- [15] C. M. Grens, *et al.*, "Large-Signal Performance, Linearity, and Reliability Characteristics of Aggressively-Biased Cascode SiGe HBTs for Power Amplifier Applications," *IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 135 - 138, September 2007.
- [16] G. Qin, *et al.*, "Linearity and dynamic load lines of CE and CB SiGe HBTs under the influence of DC bias," *International Conference on Solid-State and Integrated-Circuit Technology*, pp. 185-187, October 2006
- [17] P. R. Gray, *et al.*, Analysis and Design of Analog Integrated Circuits, 4th edition, John Wiley & Sons, 2001.
- [18] C.M. Grens, S. Seth, *et al.*, "Common Base intermodulation characteristics in advanced SiGe HBTs," *IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 244 - 247, October 2008.
- [19] C.M. Grens, S. Seth, *et al.*, "Very simple Volterra based expressions for common base intermodulation distortion in SiGe HBTs," *submitted to IEEE Transactions of Microwave Theory and Techniques*.
- [20] G. Qin, *et al.*, "SiGe HBT linearity comparison between CE and CB configurations," *Semiconductor Science and Technology*, pp. 216-220, January 2007.
- [21] G Niu, *et al.*, "RF Linearity characteristics of SiGe HBTs," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, no. 9, pp. 1558-1565, September 2001.

- [22] K. L. Fong, *et al.*, "High-frequency nonlinearity analysis of Common-Emitter and Differential pair transconductance stages," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 4, April 1998.
- [23] Q. Liang, *et al.*, "Systematic linearity analysis of RFICs using a two-port lumped-nonlinear-source model," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 5, pp. 1745-1755, May 2005.
- [24] G. Niu, *et al.*, "Characterization and modeling of intermodulation linearity in a 20 GHz SiGe HBT technology," *IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 1-4, October 2006.
- [25] M. R. Murty, *et al.*, "Implementation of a scalable and statistical VBIC model for large-signal and intermodulation distortion analysis of SiGe HBTs," *Microwave Symposium Digest, IEEE MTT-S International*, IF-TH-46, pp. 2165-2168, 2002.
- [26] M. Tutt, *et al.*, "Large Signal RF characterization of Bipolar Transistors," *IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 66-73, 2007.
- [27] S. Seth, *et al.*, "Comparing RF linearity of npn and pnp SiGe HBTs," *IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 29-32, October 2009.
- [28] H. Q. Tserng, *et al.*, "A 0.5-W complementary AlGaAs-GaAs HBT push-pull amplifier at 10 GHz," *IEEE Microwave and Guided Wave Letters*, vol. 3, pp. 45-47, February 1993.
- [29] D. Sawdai, *et al.*, "Push-pull circuits using n-p-n and p-n-p InP based HBTs for power amplification," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, no. 8, pp. 1439-1448, August 1999.
- [30] C.T. Kirk, *et al.*, "Theory of transistor cutoff frequency fall-off at high current densities," *IRE Transactions on Electron Devices*, vol. 9, pp. 164-174, March 1962.
- [31] H. F. F. Jos, *et al.*, "A model for the nonlinear base collector depletion and its influence on intermodulation distortion in bipolar transistors," *Solid-State Electronics*, vol. 33, no. 7, pp. 907-915, 1990.

- [32] H. E. Abraham, *et al.*, “Transistor design for low distortion at high frequencies,” *IEEE Transactions on Electron Devices*, vol. 23, no. 12, pp. 1290-1297, December 1976.
- [33] W.D. van Noort, *et al.*, “Reduction of UHF power transistor distortion with a nonuniform collector doping profile,” *IEEE Journal of Solid-State Circuits*, vol. 36, no. 9, pp. 1399-1407, September 2001.
- [34] M. Iwamoto, *et al.*, “Linearity characteristics of GaAs HBTs and the influence of Collector Design,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 48, no. 12, pp. 2377-2388, December 2000.
- [35] B. Gunnar Malm, *et al.*, “Ge-profile design for improved linearity of SiGe double HBTs,” *IEEE Electron Device Letters*, vol. 23, no. 1, pp. 19-21, January 2002.
- [36] J. D. Cressler, “On the potential of SiGe HBTs for extreme environment electronics,” *Proceedings of the IEEE*, vol. 93, no. 9, pp. 1559 – 1582, September 2005.